

Building on failed planetary missions

The apparent failure of a Mars probe underscores NASA's difficulties in pursuing planetary exploration. Planetary scientists must be allowed to build a credible programme of missions on the bedrock of continuing public support.

The likely fate of the Mars Polar Lander, coming on top of September's loss of the Mars Climate Orbiter, again calls into question the 'better, faster, cheaper' approach to planetary missions that the US space agency NASA has embraced under the leadership of its director, Dan Goldin (see page 565, and *Nature* **402**, 217; 1999). It also highlights the formidable technical challenges that confront attempts to probe the red planet. There have been a few successes—notably the Vikings of the 1970s, Mars Pathfinder in 1997 and the highly productive Mars Global Surveyor now in Martian orbit—but many more failures.

As another window of opportunity to fly to Mars passes without success, it may be time for NASA and the community to go back to first principles. The space agency needs to determine what must be done to meet our scientific objectives on Mars and other bodies in the Solar System, and to realistically assess how much they will cost to address.

Such an assessment may well point towards unmanned missions of greater technical sophistication—and therefore cost—than the craft that have been lost this year. For such missions to come to fruition, public opinion must be mobilized in their favour. Given the amount of attention the Polar Lander has attracted in the past week or two, that may be less daunting a challenge than it seems.

A large segment of the US public is highly motivated by the Mars missions in particular and by space exploration, manned or unmanned, in general. Most members of this group believe that it falls naturally on the US government to sponsor the exploration of the space frontier, and are happy, within reason, to pay taxes to that end.

Another sector of the public views space exploration with

apathy and even some hostility, regarding it as an extravagance irrelevant to the problems of their everyday lives. This point of view has been expanding in its reach since the end of the Apollo programme, and certainly won new adherents after the Challenger disaster in 1986. Some fear that this year's Martian mishaps will alienate even more people from the space programme. But no one has died, not much money has been spent, and there is no need for that to happen.

Advocates of a vigorous US space science programme, even if outnumbered by the apathetic and the downright antagonistic, can still have their way if they are sufficiently determined. Supporters of America's role in exploring the cosmos need to point out the obvious, budget-driven weaknesses of NASA's recent missions. The Polar Lander, for example, was built so cheaply that it lacks basic diagnostic and telemetry systems with which to pass on clues about its fate.

Proponents need to cultivate and build upon the enthusiasm of many ordinary Americans for planetary science, and they need to find articulate champions in the political arena. A prevalent theme of the current US presidential election campaign is that, after eight years of unprecedented economic success accompanied by a sense of gnawing unease about America's national purpose, voters want inspired leadership. This sentiment is being effectively tapped by the insurgent campaigns of Senator John McCain and former Senator Bill Bradley. A candidate who wants to enhance America's sense of itself could do worse than come out in support of planetary science and unmanned space exploration, pledging not to flinch from the challenge posed by the lost missions to Mars. ■

Wise progress with embryos

New guidelines should reduce the scale of ethical dilemmas in embryo research.

Given the ethical arguments that revolve around the status of the embryo, getting consensus on whether human embryo research should be allowed is impossible. Even the major religions disagree. The Roman Catholic Church states that an embryo is a person from conception, Judaism from 40 days, while Protestant churches say that it is neither a person nor a simple object. German law mirrors the Roman Catholic position and has a total ban. French law requires 'respect' for the embryo from conception, but gives legal rights only at birth; it has a *de facto* ban. Britain, Canada and Australia not only authorize embryo research, but allow embryos to be deliberately created for research purposes. So does the United States, where an absurd situation exists in which almost anything goes in the private sector, while Congress bans the use of federal funds for any such research.

The therapeutic promise of human embryonic stem cells—a new era in transplantation and cell therapy will surely emerge from their ability to divide indefinitely and differentiate into all sorts of human

tissues—obliges countries to revisit the issue of human embryo research. It has already prompted France's Conseil d'Etat, a sort of Supreme Court, and the US National Institutes of Health to do so. Both have arrived at the same conclusion: that such research should be restricted to embryos left over from *in vitro* fertilization that would be destroyed anyway (see page 565).

That is a wise and pragmatic position. Whatever one feels about the destruction of a human embryo, there is a major difference between creating embryos deliberately for research and doing research on embryos destined to be destroyed in any case. The two national bodies have usefully reduced the scale of the ethical dilemma. Moreover, which of the following is morally better, to allow spare embryos to be destroyed or to use them for research that could benefit the seriously ill? Ironically, embryo research generally, by improving the efficiency of *in vitro* fertilization and thus reducing the production of spare embryos, could avoid a future ethical dilemma posed by millions of embryos languishing in cryopreservation worldwide. ■





Biosphere 2
Columbia set to
continue running
born-again centre
p567



French genes
Controversy over
plan for access to
sequence data
p569



Fusion lobby
Europeans use spin-
off technology to
boost reactor plans
p570



'Frankenfood'
US Food and Drug
Administration
consults the public
p571

NASA facing awkward questions as hopes for Mars lander fade

Washington & Munich

Hopes were fading early this week that ground controllers will ever hear from the Mars Polar Lander (MPL) and two small penetrator probes. The craft arrived in the vicinity of Mars last Friday (3 December), and then vanished without trace.

Although last-ditch efforts to contact the robots will continue for up to two weeks, US space officials face the likelihood that all four spacecraft sent to Mars this year are lost.

The failure of the \$165 million MPL and \$30 million Deep Space 2 probes, coming after the embarrassing loss of the \$125 million Mars Climate Orbiter in September (see *Nature* **402**, 221; 1999), would be a severe blow to NASA's plan to visit Mars every two years, and may prompt a re-evaluation of the agency's entire science programme.

An accident review board similar to the one for the climate orbiter will almost certainly be set up, but there will be far less information on which to base the autopsy. Partly as an economy measure, the MPL had no means of communicating with Earth during its final descent to Mars.

With no data, engineers do not know if the probes and lander separated from one another as planned, or if the lander's novel descent engine worked properly. The terrain may have been too rough for the MPL to land upright, and the probes may have been buried in sand or lost in a steep crater.

NASA hopes to use the camera on the Mars Global Surveyor to search for the lander's shadow or its parachute. But spotting the vehicle or its debris from orbit is a long shot. NASA may therefore never learn what went wrong — a troubling prospect, as construction of a similar lander, to be launched in 2001, is already under way.

The questions are likely to go deeper than the cause of this accident. Planetary scientist Ron Greeley, of Arizona State University, says that if the MPL is declared dead there will probably be several outside reviews. Greeley was on the science team for the MPL, and chairs a committee to advise NASA on its Mars programme.

"We want to see first what the problem with the system was, and how we got to where



Anybody there? The Mars Polar Lander team wait in vain for news of their spacecraft.

we are today," he says. "We also want to see how the overall strategy might be re-examined from a scientific perspective."

NASA's philosophy of 'better, faster, cheaper' space missions is also likely to come under scrutiny. With the increasingly gloomy prospects for the MPL, a number of

scientists are questioning whether the agency's economy drive has gone too far, increasing the risk of costly mistakes.

Kristian Schlegel, a spokesman for the Max Planck Institute of Aeronomy, in Katlenburg-Lindau, Germany, which built the lander's robotic arm camera, says that "the lost opportunity for scientific results" hurts more than the money. The camera cost around DM1 million (US\$520,000).

But Schlegel does not think that there will be a loss of public confidence in unmanned space missions in Europe. Peter Wenzel, head of Solar System Sciences at the European Space Agency, believes that the agency's Mars Express mission — the first of its cheaper, flexible missions, due for launch in 2003 — will continue as planned.

Wenzel says: "We will need to look carefully at our plans to make sure that no corners are cut." The flexible European missions have lower costs "but risks are taken with responsibility," he says.

Beginning with a Soviet attempt in 1988, seven of the last ten missions to Mars have failed.

Tony Reichhardt & Alison Abbott

France to lift embryo research ban?

Paris

France's 1994 ban on human embryo research should be lifted, according to the Conseil d'Etat, one of the three arms of the country's supreme court. Its aim is to allow scientists to pursue the therapeutic promise of human embryonic stem cells.

The recommendation is one of the main conclusions of a report adopted last week by the body's assembly general. Lionel Jospin, the prime minister, commissioned the report in preparation for a parliamentary debate next year that will lead to a revision of France's bioethics legislation.

The legislation was designed so that it could be updated in response to scientific progress and shifts in moral attitudes. The report recommends that embryo research be allowed for stem-cell research under strict conditions.

Research would only be allowed on surplus embryos. There are 30,000 of these in cold storage in France, left over from *in vitro* fertilization by parents who do not wish to have them implanted. Implantation of experimental embryos to produce children would be banned.

The report also recommends allowing embryo research aimed at improving reproductive medicine. It proposes the creation of a body, akin to the UK Human Fertilization and Embryo Authority, that would approve research protocols on a case-by-case basis.

The original bioethics laws were adopted under a conservative government, dominated by Roman Catholic ideas. But the national assembly now has a Socialist majority, with a traditionally more liberal stance on embryo research.

Declan Butler

Indian patent law will protect crops 'unless injurious to health'

New Delhi

The Indian government has presented a bill to parliament proposing a law giving patent protection to new crop varieties for 15 years, but allowing patent applications to be rejected for varieties for which the commercial exploitation "may be injurious to health".

The clause allowing rejection is seen as a bid to ban both the import of genetically modified foods and the cultivation of transgenic plants where questions remain about their safety for human health or the environment.

To be eligible for patent protection, a variety "must be new, distinct, uniform and stable". But those containing "harmful genes or gene sequences like the terminator gene" will be refused registration.

Judgements about both patentability and safety will be made by a National Plant Protection Authority that will be created under the act to implement all provisions of the legislation, including testing. It will use existing lab facilities of the Indian Council of Agricultural Research until it sets up its own.

The council says the safety criteria are listed in the recombinant-DNA research guidelines issued by the Department of Biotechnology.

The legislation will not affect the right of farmers to save, exchange or sell produce of the protected variety, or the right of researchers to have access for research. But a National Gene Fund will pass some of the sale proceeds of new varieties to the village communities that preserved the plant genetic resources.

The government claims that the legislation, by ensuring a return for investment, will "facilitate the growth of the seed industry through domestic and foreign investment".

K. S. Jayaraman



EU candidate states impress with Framework proposals

London

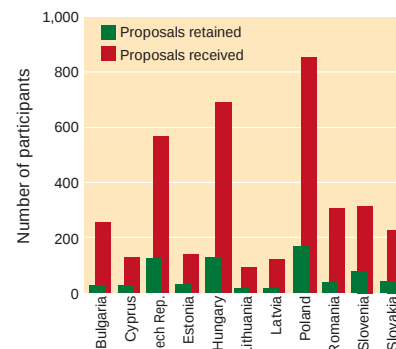
Cyprus, Estonia, Lithuania and Slovenia achieved the highest success rate of the 11 'candidate' countries seeking membership of the European Union in the first round of funding applications for the EU's fifth Framework research programme (FP5). Their achievement emerged from an analysis presented last week to a meeting of the council of research ministers in Brussels.

The European Commission was reporting on the progress of candidate countries within FP5. Research is the first sector in which candidate countries fully share the implementation of EU policy under similar conditions to member states. The commission described the results as "encouraging".

Around 12,000 proposals were made to FP5 programmes, with 29 per cent being approved for possible funding. Of the approved applicants, four per cent were from candidate countries, and, within the thematic programmes, candidate countries participated in a quarter of the proposals.

Poland, Hungary and the Czech Republic made most proposals (see graph), and proposers are mainly from the public sector. Romania performed surprisingly poorly, given its high level of participation in earlier programmes, and the reasons for this are to be investigated further.

At the same meeting, the council of ministers agreed on the timetable for the sixth Framework programme and on funding the dismantling of its nuclear installations. It also adopted a resolution on a European strategy for space, and is considering working with the



How candidate countries scored with their Framework funding proposals.

European Space Agency on projects such as Galileo and Ariane. There was also agreement to create a reserve fund to allow swift action on new research priorities. In the past, the commission has struggled to find enough flexibility in research funds to provide for areas such as AIDS and BSE.

The topic of women in science also featured prominently at the meeting, where the commission's expert report on the issue in member states was presented (see *Nature* 402, 337; 1999). Some member states called for "practical action", said a spokesperson.

The German delegation suggested that research organizations that do particularly well in recruiting women should receive a financial bonus. Such a scheme has been successful in Germany, and this proposal could well be taken forward at a conference in April and become a theme of the next meeting of EU research ministers in June.

Natasha Loder

US close to funding stem-cell work

Washington

The US National Institutes of Health (NIH) could provide funding for research on human embryonic stem cells as early as the spring if new guidelines, published by the NIH last week after widespread public consultation, are formally approved.

The guidelines try to distinguish between the derivation of the cells from embryos, which the US government is forbidden by Congress from funding, and research on the cells once they have been isolated, which can be funded (see *Nature* 397, 185; 1999). Under the proposed rules, stem cells can be taken only from spare frozen embryos after the conclusion of *in vitro* fertilization attempts.

"This is a way to ensure a separation between the creation of the embryo and the

decision to donate to research," says Lana Skirboll, director of science policy at the NIH. Physicians involved with *in vitro* fertilization will not be eligible for NIH funds for research on cells they derive themselves.

The rules are "quite reasonable", according to James Thomson, a researcher at the University of Wisconsin-Madison who, with private funding, first cultured pluripotent cells from embryos last November.

But the proposed informed-consent form may be sufficiently different from the one used by Thomson that it may not allow his cells to be used, he says. If necessary, his laboratory will culture another stem-cell line under the new guidelines. The guidelines are available at <http://www.nih.gov/news/stemcell/draftguidelines.htm>.

Paul Smaglik



Respectability beckons: Biosphere 2 is 'on the border' of being a fully functioning scientific facility.

Columbia extends support for 'mainstream' Biosphere 2

Tucson, Arizona

Columbia University in New York is poised to extend its agreement to manage the controversial Biosphere 2 centre in Arizona for a further ten years. After a rocky start, the centre is claimed to be on the threshold of becoming a recognized scientific facility.

Columbia's trustees were due to meet this week to consider university president George Rupp's recommendation to keep on running Biosphere 2 as a 'Columbia West' facility. The trustees are expected to approve the agreement, which was worked out a year earlier than had been anticipated.

Since taking over management of Biosphere 2 in 1996, Columbia has worked to turn the centre, created as an experimental, self-contained environment, into a mainstream educational and scientific facility (see *Nature* 378, 223; 1995 & 380, 275; 1996).

Increasing student enrolment confirms the success of the educational mission. But scientific success has been more elusive, because of the complexities of revamping and staffing the facility outside Tucson — which includes a saltwater lagoon, a tropical forest and carbon dioxide experiments with 50-foot-tall trees.

"It took us two years longer than we hoped, but we have turned the corner scientifically," says Michael Crow, an environmental scientist and Columbia's executive vice-provost, pointing to the budding number of publications emerging from the facility.

Last week, scientists and officials from the US Department of Energy visited Biosphere 2 as part of the centre's efforts to attract federal funding. Discussions are under way with the Los Alamos National Laboratory to conduct experiments in Biosphere 2's controlled environment. And a biodiversity research programme is being created outside the structure near Oracle, Arizona, to study the high-desert region.

"Things are moving quickly now" to meet the facility's scientific goals, says William Harris, a former assistant director at the National Science Foundation who has run Biosphere 2 since late 1996.

Joe Berry, a biologist at the Carnegie Institute at Stanford University who is one of three scientists overseeing projects at Biosphere 2, says the operation is "on the border" of being a fully functioning scientific facility "We have been doing the easy experiments," Berry says. "We have not done the hard ones yet."

Berry, who has been setting up experiments on mass balance in terrestrial ecosystems, says that implementing proper controls in the cavernous environmental lab has been extremely challenging. For example, raising Biosphere 2's internal monitoring equipment to the required standards involved major reassessments.

When Columbia took over, Biosphere 2 had about 700 sensors buried and placed around its different habitats and connected to computers. But the new team didn't know where they all were or what they did. Eventually a revised monitoring system using about 70 sensors was devised.

Throughout the four-year 'proof-of-concept' period since Columbia took over, Biosphere has continued to receive financial support from Edward Bass, the Texas billionaire who built the facility to explore his environmental visions. According to his associate, Martin Bowen, Bass is "very, very satisfied with the management and accomplishments" of Columbia.

Since Columbia took over, university records show that Bass has given Biosphere 2 about \$10 million per year (according to its tax returns, the centre has an annual operating budget of nearly \$12 million). The records say he plans to contribute \$30 million during the next five years.

About 100 undergraduate students come to Biosphere 2 from Columbia and more than 20 partner universities, paying Columbia about \$12,000 per semester. Researchers, such as Columbia's Wallace Broecker, say the courses have been very popular.

Once the new agreement is finalized, Harris says new staff will be appointed, the student population expanded to 300, and a masters' programme developed. **Rex Dalton**

US trade rules seek to balance health and patent protection

Washington & Cape Town

The US administration is relaxing its approach to developing countries seeking affordable medicines during public health crises. The move follows protests that Vice-President Al Gore and trade officials put pressure on South Africa to alter a 1997 law — aimed at combating its AIDS crisis — that allows the cheap manufacture or import of patented drugs (see *Nature* 399, 717; 1999).

Speaking at the World Trade Organization (WTO) meeting in Seattle last week, US President Bill Clinton said that US trade officials will in future consider the public health conditions of developing countries, as well as intellectual property law, in making trade decisions involving medicines.

Specifically, the office of the United States Trade Representative (USTR) will confer with the Department of Health and Human Services when a developing country complains that US trade law is making medicine unaffordable. According to a press release, USTR will "give full weight" to the department's advice on health considerations.

But the administration insists that its policy will not jeopardize the intellectual property rights of US pharmaceutical companies. Charlene Barshefsky, the US trade representative, says that she and health secretary Donna Shalala "believe that sound public health policy and intellectual property protection are, and must continue to be, mutually supportive".

US activists have welcomed the announcement. "It's a big deal," says James Love, director of the Washington-based Consumer Project on Technology. "They have reclassified these [situations] as public health issues as opposed to commercial disputes."

A US official says the United States has not changed its interpretation of a key WTO agreement on trade-related intellectual property rights (TRIPs). This allows countries to manufacture or import drugs cheaply, paying only negotiated royalties to the patent holder, only in cases of "national emergency or other circumstances of extreme urgency". South African and US trade officials had disagreed on whether South Africa's law breaches the TRIPs agreement.

The USTR said that South Africa had been removed from a 'watch list' of

countries threatened with sanctions, partly because of an agreement in which both countries promised to protect intellectual property rights while addressing the health issues.

It is not clear how South Africa could bring its legislation in line with its government's commitment to TRIPs. But Alec Erwin, the country's minister for Trade and Industry, maintains that South Africa has sought no agreement with the US on patent right exhaustion.

This occurs when the patent holder loses control over the further sale of his products, and — unlike parallel importing — is not prohibited by TRIPs. In terms of the treaty, compulsory licensing is permitted at the discretion of individual countries.

Local firms could then be licensed to manufacture goods using another company's patented formula, provided that the patent owner is compensated and the government's action is subject to judicial review. This provision had previously been opposed by the US government, which has reserved the right to pressure countries not to use it.

South Africa has not withdrawn its intention to use parallel imports or compulsory licensing to obtain drugs at lower prices, nor did the US government sanction these practices. But parallel importing remains a likely subject of proposed reform of the South African Medicines and Related Substances Act.

Meredith Wadman & Michael Cherry

AAAS to honour 'persecuted' Ukrainian marine biologist

London

Sergei Piontkovski, the Ukrainian marine biologist facing criminal charges over his role in various international research projects, is to be honoured at a reception held annually by the American Association for the Advancement of Science (AAAS) to recognize persecuted scientists.

Piontkovski and colleagues at the Institute of Biology of the Southern Seas in Sebastopol have been under close scrutiny from the local branch of the country's security services — formerly the KGB — since the end of October over their links with the West (see *Nature* 401, 835 & 402, 6; 1999).

The AAAS's Science and Human Rights Program will host the reception at the association's annual conference in Washington next February. In a letter to Piontkovski, Elisa Muñoz of the AAAS says that "the Program will honor you in recognition of the persecution that you have endured". She adds that AAAS officials "have been following your case and will continue to monitor events with concern".

Initially, Piontkovski faced charges related to espionage and hard currency concerning a number of international science grants. However, he has now been charged with illegal currency operations and organized crime in relation to his handling of project funds

paid to him in hard currency. Piontkovski expects court proceedings to start soon, and the case might start in January.

INTAS, the European body that promotes cooperation with scientists from the former Soviet Union, is one of the agencies making payments to Piontkovski, on a project on plankton distribution jointly funded by the Ukrainian Ministry of Science.

INTAS secretary David Gould said last week, "We have agreements of scientific cooperation which clearly state the terms of the contract".

Such terms include the conditions under which payment can be made, and foresee payment in hard currency. INTAS says that, as far as it is aware, currency has been handled in accordance with signed agreements.

INTAS is due to hold a meeting in mid-January to evaluate bids from newly emerging states for the 1999 funding call. "Clearly, if this is not resolved there could be consequences for the support of projects in Ukraine," says Gould. INTAS says it is continuing its "dialogue" with the science ministry.

Ukraine's president, Leonid Kuchma, has been reported as seeking a sign that Ukraine might achieve membership of the European Union. One union official says that, given the current situation in Ukraine, it will get a "negative response".

Natasha Loder

Berlin places genomics among top funding priorities

Munich

Research in Berlin that is supported by the German capital is to be focused on twelve areas — including genomics. The areas have been identified by an expert panel as reflecting the city's scientific strengths.

Ingolf Hertel, the state secretary for research in the Berlin ministry of science, has also announced plans to increase competition for funds between universities and research institutes. A small percentage of the DM3.5 billion (US\$1.8 billion) annual budget for institutional funding of science and education will be distributed competitively.

Both moves are intended to improve the quality of publicly funded research and the effectiveness with which research funds are used. Priority areas include molecular medicine and genome research, information technology, transport research, material research, optics, Earth sciences, applied mathematics and social studies.

The topics were selected by the so-called strategy forum on science, research and



Hotspot: a laboratory building of the Max Delbrück Centre for Molecular Medicine.

innovation. This was set up in January by university presidents, science ministry officials and representatives from industry and from the large science and technology centres Buch and Adlershof.

Since reunification, Berlin has become one of Germany's scientific hotspots, and now hosts three universities and 38 research institutes. But the science budget of the formerly isolated — and heavily subsidized —

city has fallen by several hundred million Deutschmarks since 1990 as subsidies from the federal government have been cut.

"Re-modelling Berlin's scientific landscape under constant budgetary constraints was, and is, a notoriously hard job," says Hertel. "Streamlining research and enlarging technology foresight will therefore be major political priorities in the next years."

But he adds that the ministry has no desire to dictate the kind of research scientists should do. "The new funding concept will be worked out jointly with the scientific community," he says.

Hans Lehrach, director of the Berlin-based Max Planck Institute for Molecular Genetics, hopes the new scheme will boost genome research in Berlin. Together with other leading genomics researchers, Lehrach has drafted a proposal for an interdisciplinary genome research centre that would bring together the various aspects of genome research carried out in Berlin. **Quirin Schiermeier**

French plan to exploit genome sparks row ...

Evry

France last week launched a FF1 billion (US\$153 million) research effort, called GenHomme, to generate economic benefits from the post-sequencing phase of the human genome project. The programme was launched by Claude Allègre, the research minister, and Christian Sautter, the economics minister, in the presence of prime minister Lionel Jospin.

But a decision to restrict the public release of annotated data — sequences in which the structure and function of genes have been identified — may put France on a collision course with its partners in the publicly funded human genome project.

The launch took place at the Genopole, France's fledgling 'genetics valley', at Evry on the southern outskirts of Paris. The Genopole will be the core of the nationwide consortium of public laboratories and private companies that is being formed to run the programme.

Members of the consortium from the private sector are likely to include the pharmaceutical companies Pasteur Mérieux Connaught, Rhône-Poulenc Rorer, and biotechnology companies, including Genset and Transgene. Public-sectors members will include the National Sequencing Centre, the National Genotyping Centre, research centres in Toulouse, Strasbourg, Montpellier and Lille, and the Pasteur Institute in Paris.

According to Alain Henault, Allègre's technical adviser for genome affairs, the consortium will be open to all, including foreign companies, if they comply with a charter governing the terms of collaboration.

Five of the charter's six clauses concern confidentiality. One says that annotated sequences cannot be submitted to public databases before they have been subject to intellectual property protection. Another forbids participants from communicating results "either orally or in writing" without permission from their partners, who will have 60 days to decide whether to allow or delay publication.

Allègre said last week that GenHomme "clarified the position on intellectual property from the outset," arguing that this was "important, to avoid controversy later". But he may have been over-optimistic. Shown a copy of the charter, Jean Weissenbach, the head of the National Sequencing Centre, who has complained to Jospin about US companies holding back genome data, said that he had not seen it before and would refuse to sign it.

Pierre Tambourin, the head of the Genopole, also claimed ignorance of the charter's details. He said he was "not in agreement" with its restrictions on submissions to public databases. But Jacques Demaille, director of



Helping hand: Lionel Jospin meets sufferers from genetic diseases at the launch of GenHomme.

the science ministry's genome activities, says that the charter is not open to negotiation. "Weissenbach is under no obligation to sign it," he says — adding that, if he refuses, he would be unable to participate in GenHomme projects.

Henault says that "collaboration with the private sector is impossible without such confidentiality clauses". But France's robust approach to intellectual property seems likely to reopen the international debate over access to sequence data.

Axel Kahn, a prominent French geneti-

cist, who is also deputy scientific director for health and agricultural biotechnology at Rhône-Poulenc, says that protecting annotated data is acceptable in an industrial collaboration, for example on a particular gene. But "it is highly contestable, if we are talking about protection of sequences that are publicly available and have been annotated 'in silico' [by computer]".

The Bermuda agreement, which requires that all genome data be made public within 24 hours, only covers raw sequence data. But David Bentley, head of human genetics at the Sanger Centre at Cambridge in the United Kingdom, says: "I would strongly argue in favour of the open and free release of interpretation and annotation that leads to a better understanding and a more accessible genome".

Bentley says he would "welcome clarification" from the French government, as to what it will patent. Such clarification will also be needed "to see how this would impact on the international collaborative scene after the sequence is completed".

In 1997, Germany withdrew plans to give industry privileged access to sequence data from its national programme, after the United States, the United Kingdom and France threatened to withhold their data from German scientists and companies (see *Nature* **387**, 536; 1997).

Declan Butler

... while Japanese sequencers feel neglected

Tokyo

Leading Japanese researchers have raised doubts about the management of the country's genome research. The warning has come despite the celebratory mood surrounding last week's announcement of the sequence of chromosome 22, in which researchers at Keio University School of Medicine played a key role (see *Nature* **402**, 447; 1999).

Speaking at a press conference in Tokyo to announce the achievement, Nobuyoshi Shimizu, Japanese partner of the chromosome 22 team, said that the government is concentrating too heavily on post-sequencing research, at the expense of genome sequencing itself.

Shimizu, professor of molecular biology at Keio, says the government is "attempting to skip sequencing" to focus on functional genomics. The shift to

post-sequencing research, including the analysis of single-nucleotide polymorphisms and full-length complementary DNA, reflects Japan's concern over its limited sequencing capacity (see *Nature* **399**, 96; 1999).

While acknowledging the significance of such research, Shimizu stresses the importance of support for sequencing the human genome. "Given Japan's limited sequencing capacity, there is a feeling within the government that further investment [in sequencing research] would not be worth the effort," he says.

Kenichi Matsubara, deputy director of the International Institute for Advanced Sciences and a pioneer of Japan's human genome project, blames the country's poor track record in genome sequencing on inefficiencies in the government funding system and its failure to

promote genome research in a coordinated way.

According to Matsubara, the current system, under which different ministries carry out individual genome projects, has wasted money and made the allocation of funding unclear. "By the time the money is split among five different ministries, there seems to be very little left for actual research," he said at last week's press conference.

Despite the launch of a genome-research centre at the Institute of Physical and Chemical Research (RIKEN) last year, Shimizu says this has caused the allocation of funding to become biased. Although RIKEN and Keio University are sequencing 20 per cent of chromosome 21, "the money has all gone to RIKEN, and we have to make do with just seven sequencers and no funding increase," he says. Asako Saegusa

Japan pulls its fusion research together ...

Tokyo

Japan is to link ten nuclear-fusion research institutes into a single network to boost collaboration between university research and government-run institutes.

The move is being seen as a bid to increase Japan's chance of hosting the International Thermonuclear Experimental Reactor (ITER), the proposed next-generation fusion facility. Japan is widely considered to be the most likely candidate to host ITER, as it is expected to provide the largest share of the construction cost.

The new network is the product of the Science Council of the Ministry of Education, Science, Sports and Culture (Monbusho), and links previously independent efforts in fusion research at institutes run by Monbusho and the Science and Technology Agency (STA).

The move brings together senior researchers and directors of institutes such as Monbusho's National Institute for Fusion Science (NIFS) and STA's Japanese Atomic Energy Research Institute (JAERI), and will launch research plans in seven areas. These include the development of new material for reactor-vessel walls, the creation of a database for superconductor technology and the development of novel laser technology.

The collaboration also involves the Japan Nuclear Cycle Development Institute and the universities of Kyoto, Tokyo, Okayama, Tohoku, Tsukuba, Kyushu and Osaka. It is seen as a necessary step towards the impending merger of STA and Monbusho in 2001, and the planned transformation of national universities and research institutes into 'agencies' with greater administrative independence (see *Nature* 401, 416; 1999).

Osamu Motojima, professor of plasma research at NIFS, says that Japan's fusion programmes have been carried out as isolated efforts because of bureaucratic barriers between ministries. "The merger of STA and Monbusho should help break down the barriers and allow efficient collaboration between government institutes and universities, although the merger of institutes themselves (JAERI and NIFS) should be approached more carefully," says Motojima.

"A strong research base is necessary for Japan's future role in ITER, and it is therefore essential to build a stronger link between current domestic fusion programmes," says Masami Nakamura, director of STA's nuclear fusion development office.

Hiroshi Kishimoto, an executive director of JAERI, says the network will contribute to 'satellite research' for validating ITER design, for example the development of a heat exchanger for converting the energy in the high-energy neutrons formed in the fusion process.

But he emphasizes that the planned network for ITER would have to take a more technological approach than the Monbusho-led network, which is focused mainly on academic research.

Despite US withdrawal from the ITER programme, and waning enthusiasm among other partners, Japan — which last year suggested plans for a scaled-down version of the reactor to meet budget constraints — is still keen to host the facility (*Nature* 394, 3; 1998).

But some researchers are concerned at the financial implications of Japan's enthusiasm for hosting ITER. "The reactor is not expected to begin its operation until the middle of the next decade, and we don't even know whether it will work," says one leading nuclear physicist. "We need to think carefully before making a commitment that could not only jeopardize basic research in fusion science but also worsen the nation's financial situation."

Asako Saegusa

... as Europeans lobby for reactor construction

Munich

European proponents of the International Thermonuclear Experimental Reactor (ITER) have launched a campaign to win support for the construction of the fusion reactor in the run-up to key budget decisions due next year.

The first of a series of public seminars took place in Munich last month, introducing the concept of a slimmed-down version of ITER intended to meet criticism of excessive cost. Last week ITER scientists held an exhibition at the European Parliament and lobbied members of the parliament, many of whom have expressed scepticism of fusion as a viable energy source.

The lobbying comes at a time when the ITER collaboration — a partnership between Europe, Japan and Russia — is preparing a detailed design outline for the reactor before the opening in Japan next spring of budgetary discussions for 2001.

The design will also form the basis of a proposal for funding under the European Commission's sixth Framework programme of research (FP6), due to start in 2003. Discussions about the content of FP6 begin next summer.

Plasma physicists in the collaboration believe that the engineering design phase for ITER should not be further extended. It has already been extended for an additional three years to cope with cash shortages resulting from waning political confidence (see *Nature* 387, 746; 1997). The physicists believe the time has come to bite the bullet and start building, to test the validity of



Street demo: technology developed from fusion research is being used in the energy-efficient Altrabus in Bologna, Italy.

experimental plasma-physics predictions.

"Plasma-physics research will lose its dynamism without a close coupling to energy production," Umberto Finzi, director of the European Commission's energy programme, told the Munich meeting.

The ITER programme suffered a major blow earlier this year when the United States withdrew support. The remaining three partners are now promoting a concept, known informally as ITER-Lite, which reduces technical objectives and cost.

The cost of ITER-Lite, around US\$3 billion, is less than half that of the original proposal, but the original goal of achieving ignition has been dropped. Ignition is an important phenomenon for physicists. But, to move confidently to the next step of a demonstration fusion power plant, ITER has only to demonstrate gain — energy amplification.

Both Japan and Canada (on behalf of the European partner) are hoping to offer sites to host ITER during the next couple of

years. Klaus Pinkau, recently retired director of the Max Planck Institute for Plasma Physics in Garching, told the Munich meeting he was optimistic that, once a site had been established, the United States would consider rejoining. Pinkau is co-chairman of a working group set up to oversee the design of a cheaper ITER.

A site in Canada would probably be more attractive to the United States because its geographical proximity would favour participation of US industry in the construction.

But the United States will need considerable persuasion to change its position. It argues that it does not need to look for a new, cheap source of power because it can rely on its own oil supplies.

It also has a substantial commitment to research into inertial confinement fusion, an approach that competes with ITER's magnetic confinement approach. But Pinkau said that the United States should recognize the political importance of ensuring stable energy supplies in other countries.

Alison Abbott

US food-safety body hears protests over genetically modified food

The US Food and Drug Administration is holding public meetings to listen to critics of genetically modified food. Some fear the move may backfire.

Washington

The US Food and Drug Administration (FDA), faced with increasingly vocal consumer protests about its regulation of genetically modified (GM) food, is undertaking an unprecedented programme of public consultations aimed at reassuring its critics.

But many critics remain unconvinced, and it is not clear whether the consultations will relieve public concern or intensify it.

The FDA's good name with the US public is widely regarded as a major factor in the rapid and, until this year, unobtrusive arrival of genetically modified soy and maize into the US food chain. But US environmental activists and consumer groups have recently been working hard to cultivate the kind of public backlash that has blocked GM foods in some European countries.

In October, the FDA responded by announcing that it would seek to "engage the public about foods using bioengineering" at large public meetings in Chicago on 18 November, in Washington on 30 November, and in Oakland, California, on 13 December.

Attempting to relieve public criticism through high-profile public debate was always going to be a risky strategy, however, as the first two meetings have shown. Even as they applaud the openness of the process, the FDA's critics have made full use of it to intensify their attacks on its integrity.

Inviting criticism

The format of the meetings has been criticized by some participants because public comment — from dozens of people, given two minutes each — took place late in the day, after FDA officials and discussion panels had spoken, and when most television cameras and reporters had left.

However, the FDA has ensured that critics of biotechnology are represented on its discussion panels. At the Washington meeting, for example, half of a six-strong panel discussing 'science, safety and regulatory issues' were generally opposed to the use of GM foods. One, Steven Druker of the Alliance for Bio-Integrity, is suing the FDA to demand labelling and more safety testing of GM foods.

"I'm surprised to be here, and the fact that I am suggests that the FDA is open to all sides



Made for TV: the meeting in Chicago provided an opportunity for demonstrators to make their point on television.

of the argument, and may be open to change," Druker said at the panel discussion, which was dominated by attacks on the FDA's position.

For example, Carol Foreman, director of food policy at the Consumer Federation of America, charged that the FDA's current voluntary policy for regulating GM foods was developed in 1992 "as part of the regulatory relief programme" of president George Bush.

James Maryanski, the FDA's food biotechnology coordinator, hotly denies this. "I can assure you that our policy was developed by FDA scientists," he says. But although Maryanski says he remains open to any information that emerges at the final meeting and from written comments that the agency has also invited, he says that, so far, "we haven't heard anything that would call for a substantial change in our scientific approach" to the regulation of GM foods.

Joseph Mendelson of the Center for Food Safety thinks that the most likely outcome of the consultation process will be that the FDA will make mandatory the existing voluntary system for regulating GM foods. Under this, industry and FDA scientists confer to identify any special cause for concern about a proposed genetic modification. Such action won't satisfy the critics, who want far tighter regulation based on extensive safety testing for each genetic modification.

Scientists involved in the panel discussions, such as Peter Day, head of the biotechnology centre at Rutgers University, New Jersey, applauded the FDA's process. "I admire what they are doing," he says. "If you are going to be credible, you've got to include the whole spectrum of opinion."

Extensive coverage

Industry representatives had called for the FDA to explain itself to the public, and those who spoke at the Washington meeting also lauded the process. But it is by no means clear that industry will benefit from it.

Extensive television and newspaper coverage of both the Chicago and Washington meetings tended to reflect the colourful anti-GM demonstrators outside and the equally colourful anti-GM arguments within. The latter ranged from papers, produced by Druker, showing that the 1992 FDA regulations were subject to fierce internal debate, and allegations from the floor that cows are refusing to eat GM maize feed.

"You've got to wonder why the FDA is doing this, when it provides an avenue for them to take such a beating," reflects Mendelson. Part of the answer is that the GM food debate is providing a stiff and unusually sophisticated challenge to the agency's credibility. Agency officials hope that their response will be sufficiently open, and convincing, to maintain the US public's faith in GM foods — and in the FDA. Colin Macilwain

Innovative projects favoured as China ups research funds

Beijing China's major research funding agency, the Natural Science Foundation of China (NSFC), announced last week that it will award RMB1.08 billion (US\$130 million) to fund research approved in 1999, about RMB200 million more than last year.

The NSFC received 2,000 more applications this year than last year, 20,925 of them in the 'general projects' category. But only 16.6 per cent of applications in this category were approved in 1999, compared with 19 per cent last year. Each project received an average of RMB135,000, up 9.8 per cent from 1998.

An NSFC spokesman said special attention had been given to innovative projects, even if they were not attractive to all reviewers. Interdisciplinary research funds have been increased to RMB10 million.

MIT will export its media know-how to Ireland

London The Massachusetts Institute of Technology (MIT) and the Republic of Ireland have confirmed that they are to set up a European base of the MIT MediaLab. MediaLab

Europe will get an influx of know-how from faculty members, research and students from the MIT MediaLab. For the first ten years, both labs will share intellectual property (see *Nature* 401, 836; 1999).

The new lab will initially focus on Internet technologies and applications, and on interactive and multimedia applications. It is expected to develop a portfolio of research, applications and computational methods. It will be located in a new 'multimedia village' in Dublin. Running costs for the first ten years are estimated at \$166 million.

Crisis meeting over mouse archive's future

Munich European Union research commissioner Philippe Busquin is to call a meeting of national research councils to discuss plans to rescue the European Mouse Mutant Archive. The archive, near Rome, was thrown into crisis when the European Commission refused a request to continue contributing to its running costs (see *Nature* 402, 4; 1999).

Busquin has called for a "European solution" to the problem, but has not indicated that the commission could play a direct role. European research ministers are playing a major role in a short-term rescue plan for the European Bioinformatics Institute, an outstation of the European

Molecular Biology Laboratory that recently found itself in the same situation as the mouse archive (see *Nature* 402, 450; 1999).

Rival claim to Livermore's radar patents rejected

San Diego The US Patent and Trademark Office issued a notice last week upholding all the original patent claims of a former Lawrence Livermore National Laboratory (LLNL) scientist who invented a valuable micro-impulse radar technology. The notice rejects the competing claims of a small Alabama firm, Time Domain Inc.

The company had challenged the invention of the former LLNL scientist, Thomas McEwan (see *Nature* 400, 6; 1999). McEwan said he was pleased by the decision after a "vicious" campaign to "smear" him. Despite losing all claims to LLNL's technology, Time Domain officials said they considered the decision favourable, as it defined the LLNL's patent rights.

Telescope launch is insured against loss

London Final preparations are being made for the space launch on Ariane 5 tomorrow (10 December) of the world's largest X-ray telescope, XMM (X-ray Multi-Mirror). The

telescope is expected to cost the European Space Agency around £500 million (US\$808 million) over its lifetime.

The agency thinks XMM will be substantially under budget, which has allowed the telescope's instruments to be insured. This unusual step has been taken because of the launcher's poor track record. Project scientists said last week that problems besetting NASA's X-ray telescope, Chandra, will have minimal impact on XMM (see *Nature* **397**, 462; 1999).

Congress issues warning on spallation neutron source

Washington The Spallation Neutron Source project is again under fire from critics in the US Congress after the Department of Energy arranged to release \$50 million to begin its construction at the Oak Ridge National Laboratory, Tennessee.

James Sensenbrenner (Republican, Wisconsin), chair of the House Science Committee, and Ron Packard (Republican, California), chair of the subcommittee that funds the energy department, wrote to energy secretary Bill Richardson on 24 November, warning that release of the money defied the intent of Congress and "will only serve to weaken congressional support" for the neutron source.

DFG head moves on to Humboldt Foundation

Munich Joschka Fischer, Germany's foreign affairs minister, has appointed Wolfgang Frühwald as president of the Alexander von Humboldt Foundation, which awards grants for foreign scholars to carry out research in Germany. Frühwald, 64, is the first social scientist to head the foundation. He is a professor of German literature and former president of the university research funding

agency, the Deutsche Forschungsgemeinschaft.

Frühwald takes over from astrophysicist Reimar Lüst, 76, former director-general of the European Space Agency and former president of the Max Planck Society. Frühwald is the fifth president of the Humboldt Foundation since it was re-founded after the Second World War in 1953. Since then, the foundation has supported more than 18,000 highly qualified research fellows from 123 countries.

Lexicon Genetics

The 7 October issue of *Nature* carried a News story by Rex Dalton describing a number of researchers' difficulties in getting hold of materials presented in a paper published in *Nature* on 9 April 1998 (see *Nature* **392**, 608–611; 1998 & **401**, 515–516; 1999).

Having investigated these problems, we have concluded that Lexicon has not in practice contravened *Nature's* policy on the free supply of materials. Consequently, we do not intend to impose any sanctions on the authors.

Researchers were understandably frustrated by lengthy delays in the supply of materials and a lack of clarity in communication. A major contributor to these problems was the material transfer agreement (MTA) — negotiations between Lexicon and the researchers' funding or host institution were complex and protracted.

In this instance, Lexicon's MTA was drawn up with input from, and full disclosure to, *Nature* editorial staff. It was our view that the MTA struck a balance between access to materials for other researchers and protection of the investment made by Lexicon. Clearly this has not worked well in practice. The conflicts described in our News story are a growing feature of interactions between academic and industrial researchers and deserve high-level attention.

Lexicon has reaffirmed its willingness to supply materials described in its *Nature* paper in accordance with the MTA. We encourage readers who have comments on this process, or more generally on the question of the extent to which MTAs can be a barrier to research progress, to contact us.

Philip Campbell
Editor

Seeking clarity in the debate over the safety of GM foods

Sir—A more rigorous analysis of the debate over GM versus non-GM crops is now needed. A fundamental flaw in most arguments (for example, the debate on substantial equivalence^{1,2}) is the erroneous treatment of all genetically modified (GM) plants as one homogeneous group. This distorts arguments and weakens reasoning. It is time to introduce a classification of the distinctive types of GM crops. I propose the use of three classes:

- (1) 'Wide transfer', referring to the movement of genes from organisms of other kingdoms into plants;
- (2) 'Close transfer', referring to movements of genes between species of plants; and
- (3) 'Tweaking', referring to the manipulation of levels or patterns of expression of genes already present in a plant's genome.

Genetic modification, in its strictest meaning, has been performed for 10,000 years, including notably gross manipulations such as the development of hexaploid wheat and triticale. Consequently, I think it would also be sensible (and politically appropriate!) to refer to GM crops generated using modern biotechnology as 'new GM crops', as distinct from the 'old GM' of traditional breeding technologies.

Use of such terminology will throw into relief some of the more extreme claims made by both sides of the argument. For example, the use of 'substantial equivalence' (as employed for conventional crops) is most likely to be sufficient for crops developed using 'close transfer' and 'tweaking', as the results of these manipulations are unlikely to be different from processes used by traditional breeders. However, the introduction into the food chain of significant amounts of novel proteins by 'wide transfer' may well require more thorough testing, as for the introduction of a drug. If such tests were introduced only for plants generated by wide transfer, then the obvious objections to this suggestion raised by Anthony Trewavas and C. J. Leaver² would not apply.

Use of a refined categorization of the new GM crops would focus arguments and facilitate more balanced conclusions in a currently unnecessarily polarized debate.

Mark Tester

Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK

Sir—Millstone *et al.* in their ill-informed Commentary "Beyond 'substantial equivalence'"¹ advocate the use of extensive animal toxicology tests on whole

GM foods. But this type of testing has never been used to evaluate the safety of traditional, non-GM foods even though they often contain hundreds of unique proteins and other constituents. Such testing of whole foods would be tremendously unfocused, wasteful of laboratory animals, and unlikely to detect harmful substances, even if they were present.

The novel proteins in glyphosate-resistant soya beans are present at very low levels, for example, and their effects, if any, would not be detected by feeding whole soya beans to lab animals. A focused approach aimed at these novel proteins identified through substantial equivalence would be much more reasonable. This is the approach that has been used and it clearly demonstrates the safety of these novel components. The rest of the soya bean is identical (or substantially equivalent) to traditional soya beans and, in our view, is safe for consumers.

Millstone *et al.* also suggest that glyphosate-resistant soya beans exposed to glyphosate may have different composition from traditional soya beans. In our view, such concerns are not particularly valid. The composition of specific components in specific crops, GM or non-GM, varies within a range as a result of variety, environmental/climatic conditions and agricultural practices. Soya beans exposed to glyphosate are not altered in composition beyond the normal range that exists naturally. And, even if they were, animal toxicology testing would be unlikely to reveal such differences.

Steve L. Taylor*, Susan L. Hefley†

**Department of Food Science and Technology, 143 Food Industry Complex, Lincoln, New England 68583-0919, USA*

†Food Allergy Research & Resource Program, Department of Food Science and Technology, 351 Food Industry Complex, Lincoln, New England 68583-0919, USA

Sir—One reason for the unsatisfactory state of GM food regulation¹ is the assumption, stated in the FAO/WHO biotechnology and food safety report³ and adopted by UK regulators, that genetic engineering does not differ from conventional selective breeding. Hence, Trewavas and Leaver² conclude that GM food should not be subject to more rigorous testing than novel non-GM food.

I believe this assumption is erroneous. Genetic engineering enables exotic genes from viruses and bacteria and other non-food species to be introduced into food crops. These genes are combined in novel constructs, often with viral promoters to make genes overexpress continuously. The constructs are inserted into genomes by transformation techniques that cannot control where the genes go, resulting in a

range of unpredictable positional effects and rearrangements. I would prefer to have a moratorium on environmental releases of GM material until safety issues have been more adequately and openly addressed.

Mae-Wan Ho

Institute of Science in Society and Department of Biological Sciences, Open University, Walton Hall, Milton Keynes MK7 6AA, UK

Millstone et al. reply—Our Commentary¹ on the use of the concept of substantial equivalence in evaluating the safety of GM foods has triggered a lively debate, but several correspondents have criticized us for things that we did not say, and in our view none of them has addressed our central argument that using substantial equivalence discourages vital scientific investigation.

Trewavas and Leaver² and M. J. Gasson⁴ have suggested that we argued for full toxicity testing of all novel crop seeds in all potential circumstances of use. We did not make that assertion, but rather looked askance at an approval system that, if measured against the yardsticks of scientific rigour and public confidence, we believe is not adequate or effective. We also sought to unravel some of the reasons why confidence is lacking, and suggested ways in which the approvals system might be strengthened.

We do not agree with the view that single-gene changes necessarily result in well-characterized plant responses. For example, experimental genetic manipulation of oilseed crops, including rape, has led to the unexpected discovery of enzymes and regulatory mechanisms affecting lipid metabolism. Gasson's confidence that unexpected alterations in crop chemistry as a result of genetic modification are unlikely is contradicted by a report⁵ at a recent meeting that glyphosate tolerance in Roundup Ready soya beans appeared to be occurring at the expense of diminished heat tolerance as a result of changes to the lignin content. That research was triggered by the unexpectedly poor performance of Roundup Ready soya bean in hot climates, and it was only biological testing under various conditions that revealed the effect.

We totally reject the accusations made by Derek Burke⁶. Our Commentary contained no conspiracy allegation, nor did we criticize any individual—merely 'wishful thinking'. There seems to be an unwillingness on the part of some enthusiasts for GM foods to engage with scientific challenges and policy questions such as those we posed. We are not demanding exhaustive toxicity tests. Rather, we hope to encourage a re-examination of what constitutes an adequate risk assessment for GM and other novel foods. This might help to improve public acceptance of the technology.

Erik Millstone*, Eric Brunner†, Sue Mayer‡

**SPRU—Science and Technology Policy Research,*

Mantell Building, University of Sussex, Brighton
BN1 9RF, UK

†Department of Epidemiology and Public Health,
University College London, London WC1E 6BT, UK

‡GeneWatch UK, The Courtyard, Whitecross Road,
Tideswell, Buxton, Derbyshire SK17 8NY, UK

1. Millstone, E., Brunner, E. & Mayer, S. *Nature* **401**, 525–526 (1999).
2. Trevas, A. & Leaver, C. J. *Nature* **401**, 640 (1999).
3. Food & Agriculture Organization/World Health Organization
*Biotechnology and Food Safety: Report of a Joint FAO/WHO
Consultation* (FAO, Rome, 1997).
4. Gasson, M. J. *Nature* **402**, 229 (1999).
5. Getz, J. M., Venecil, W. K. & Hill, N. S. *The 1999 Brighton
Conference* **8C-6**, 835–840 (British Crop Protection Council,
Farnham, 1999).
6. Burke, D. *Nature* **401**, 640–641 (1999).

Why Diamond should go to Daresbury

Sir—You report that the Wellcome Trust favours siting the new Anglo-French synchrotron source (Diamond) at the Rutherford Appleton Laboratory in southern England instead of at Daresbury Laboratory near Manchester (*Nature* **402**, 451; 1999). I am concerned about Wellcome's statement that it is important to site Diamond close to the existing neutron spallation source (ISIS).

In fact, no protein-crystallography research is done on ISIS. Neutron protein-crystallography research is and will be conducted at the Institut Laue Langevin neutron source in Grenoble, France. I hold a Wellcome Trust research grant to work on both synchrotron X-ray protein-crystallography data collection and neutron protein-crystallography data collection. It has not been necessary to have X-ray and neutron sources at the same location—it is simply not a requirement.

I am surprised that I have not been asked my opinion by anyone on the question of the siting of Diamond. Not only am I the chairman of the Institut Laue Langevin neutron beam-time biology review committee, but I have also chaired synchrotron X-ray advisory committees for the European synchrotron radiation facility in France, Cornell in the United States and the Daresbury Synchrotron Radiation Source protein-crystallography beam panel.

I strongly support Daresbury as the site of Diamond. This vital project requires the vast body of experience among the staff at Daresbury Laboratory. The Rutherford Appleton Laboratory has its own expertise and is planning its own ISIS neutron source upgrades. But this is irrelevant to placing Diamond in the best UK location for its optimal, proper and successful development, which should be at Daresbury.

Finally, I warmly welcome the Wellcome Trust's involvement to the fullest extent possible in Diamond at Daresbury. The

protein-crystallography programme on Diamond must be as vibrant as possible, given the importance of structural biology in the expanding fields of structure-based drug design and health care. The trust and the UK government research agencies are key players in this enterprise that should ensure the continuing international competitiveness of UK research and its pharmaceutical industry.

John R. Helliwell

Department of Chemistry, University of Manchester,
Manchester M13 9PL, UK

Seeking a guide to the quality of web writing

Sir—The Opinion article "Dangers of over-dependence on peer-reviewed publication" presents an interesting perspective on an important debate (*Nature* **401**, 727; 1999). It refers to a study on the problem of 'defining a publication' carried out by an international ad-hoc group under our joint chairmanship at the request of the science, technical and medical publishers, and presented to their recent meeting in Frankfurt. Far from blurring the concept of a publication, our proposals attempt to impose some order on the continuum of publication that has become possible in the electronic environment.

We acknowledge the variety of forms of scientific writing found on the Internet, and offer criteria for distinguishing among them, thereby giving readers a clearer basis than now exists for evaluating the status and credibility of whatever they are reading. We invite your readers to examine our discussion document (at the website <http://associnst.ox.ac.uk/~icsuinfo/aaas-stm.htm>) and to contribute to the refinement or rebuttal of these ideas.

Roger Elliott*, **Mark S. Frankel†**

*Chairman, ICSU Press,

Department of Theoretical Physics, University of
Oxford, 1 Keble Road, Oxford OX1 3NP, UK

†American Association for the Advancement of
Science, 1200 New York Avenue NW, Washington
DC 20005, USA

How ethical principles can aid research

Sir—Søren Holm and John Harris argue that the precautionary principle stifles discovery (*Nature* **400**, 398; 1999). I believe such a judgement to rely on a misunderstanding of the link between science and progress.

Discovery is by definition never self-evident: when a problem arises, the most

apparently ready-made ways to solve it are likely to be the least productive in genuine novelty. This is precisely where moral issues come in, in science as in life in general. Real progress originates from the refusal to take a path that would threaten one's own moral choices and values. It is in such a situation, which to me is the essence of science as a daughter of ethics, that other ways do show up, not previously thought of, which lead to discovery. It is precisely the moral problem coming to the researcher's consciousness that alerts him to the fact that he is able to find another way. Not listening to the voice of consciousness in such a situation thus leads to missed discoveries. Therefore it is precisely in research that ethical principles are most useful.

So the precautionary principle should be viewed rather as a guide to avoid wrong directions, opening the way to better ones.

Joël Sternheimer

Reseau Associatif de Chercheurs Independants,
1 rue Descartes, 75005 Paris, France

Lobbying groups must be trustworthy

Sir—I agree with H. Mohr that Greenpeace has no credibility when it comes to evaluating science (*Nature* **401**, 524; 1999). Many of its activities and campaigns are based on propaganda rather than facts.

Academic research requires freedom from commercial competition and dependence. Mohr says he has enjoyed such freedom all his life, but this begs the question of whether this freedom is likely to be reduced if commercial interests arise. In cases where scientific results are open to different interpretations, scientists should not have a monopoly on opinion and answers, particularly on issues of public importance. Organizations representing the general public must have a say.

But these organizations must be trustworthy, rather than manipulating facts to suit their own agendas. People usually trust relief organizations, such as the Red Cross or Médecins sans Frontières, but these organizations generally avoid controversial issues. There are, however, numerous local organizations that are concerned with science in respect of both the environment and the production of food and other consumer goods. One example is Bellona in Oslo, recognized for its work with nuclear waste in the former Soviet Union. We need more of these types of organization at the international level, assuming they use acceptable methods.

Ketil Haarstad

Jordforsk, Norwegian Centre for Soil and
Environmental Research, N-1432 Ås, Norway

Expressing a consensus on candour

If governments are to define deception as research misconduct, science deserves clarity and rigour in the definition.

Louis M. Guenin

Is there a problem about misconduct in research? When this question first received widespread attention in the United States a decade ago, many would have understood it as a query about the frequency of misdeeds. As university and government investigations have since become routine, some of the public's fears about the incidence of infractions have been eased. But it happens that governments have not settled upon a rigorous definition of wrongdoing.

What is "misconduct"? A definition recently published for comment by the White House Office of Science and Technology Policy (OSTP) has returned this question to centre stage. On the premise that science deserves the best of scientific, moral and legal thinking combined, I propose here a different definition — one that not only avoids counter-examples to which OSTP's well-intentioned proposal succumbs, but that sounds, in the most fundamental notions of right and wrong, pertinent to the experimental setting.

The imagined drama

As some would have it, OSTP opens a new act in a drama featuring science as agonist, law as antagonist. The plot begins about ten years ago as aliens invade Scientia. The aliens are legislators, lawyers and bureaucrats whose avowed mission is to expel dissemblers. Scientists immediately protest that prevaricators are rare and that Scientia deserves home rule. To allay native fears, the aliens announce a "collegial" investigatory process. But the natives soon learn that under this scheme, the alien prosecutor is also the judge, and the judge conducts no trials.

In act II, some convicted scientists obtain hearings before a higher alien authority. To everyone's surprise, the reviewing tribunal acquits one accused after another. Few read the tribunal's lengthy opinions, but reports put down the acquittals to prosecutorial incompetence. Taking a longer view, native observers whimsically dubbed "fraudians" doubt whether an alien tribunal can understand Scientia. In act III, frustrated alien prosecutors attempt to dragoon the natives into fewer hearings. Meanwhile alien leaders decide to solicit native views about misconduct. Fraudians organize a competition to select that acceptable definition of misconduct containing the fewest alien legalisms. Many committees submit brief definitions, but each entry is either too vague or succumbs to counter-examples, and none gains

acceptance in Scientia. As act IV opens, the fraudians have scanned OSTP's entry and awarded it an impressively low legalism score. But soon counter-examples appear. In the definition derby, there has yet to be a winner.

This drama accentuates the influence of a perception of science and law at loggerheads. To be sure, the law often mangles the very notion of science. As we encounter the US Supreme Court defining science, we can only despair to read one account (Sir Karl Popper's) unwittingly adopted as if there were no rival¹. But in respect of misconduct cases, two observations are telling. First, the seldom-read opinions of the government's reviewing tribunal reveal earnest study of pertinent science, cogent reasoning and painstaking appraisal of conflicting evidence. Opinions in the most publicized cases — like Thereza Imanishi-Kari's, almost all acquittals — encourage the belief that this panel can capably adjudicate charges of wrongdoing². Many more cases have ended in admissions of guilt.

Second, in the throes of scruples about language, it has escaped notice that a consensus obtains among scientists about what conduct is misconduct. Consider how rarely scientists disagree, in respect of untruthfulness, about a statement of the form "Act *aby* *S* was wrong". Fraudians have been unable to find any category of conduct that one thinks wrong and another innocent. But by dint of antixenophobia — avoidance of a foreign tongue — the definition derby has yielded only imprecise sketches of what constitutes misconduct.

By unshackling the linguistic tools with which we have learnt to express norms of conduct, we might capture this scientific consensus. I take as my task to bag that quarry. A search for norms can be understood as an exercise in hypothesis formation. The data consist of our judgements about what is right and wrong. We test a hypothesis to see if it explains and correctly predicts those judgements. If a draft norm condemns an action that a scientific consensus absolves, or vice versa, we reject the norm, draft another, and test the new one.

Absence of a clear definition has impeded effective oversight and policies.

Groping for norms of candour

"Integrity", despite early institutionalization in "Office of Research Integrity" (ORI), is not the root notion here. Integrity is variously understood as a set of virtuous dispositions or, after *integer* or "whole", as the attribute of so adhering to a set of commitments as to keep oneself intact. The pivotal concept here is candour, the attribute on a given occasion of not uttering anything that one believes false or misleading. We describe breaches of candour as deception. The government might reasonably assert that one who stoops to deception in quest of distinction betrays such lack of distinction that further support would waste public funds³. An incident of self-promoting deception might also be viewed as a selective disadvantage reliably predicting failure in grant competition. Since 1987 the US government, followed recently by European bodies, has prohibited "fabrication, falsification, and plagiarism". Despite the non-moral senses of "fabrication" (as synonym of "construction") and "falsification" (as antonym of "verification"), the regulations toss off the foregoing offences without definition. They then add another, "serious deviation" from "accepted practices".

A National Academy of Sciences panel observed early on that "absence of a clear, explicit definition ... has impeded development of effective institutional oversight and government policies and procedures"⁴. Although we inherit a storehouse of philosophical and legal machinery bearing on breaches of candour, two reasons have been urged for shunning it. First, it is suggested that such machinery will not work for the laboratory and that scientists would not readily comprehend it anyway. In fact, scientists have criticized the sketchy misconduct texts of the past decade as vague and overreaching; misrepresentation of material facts, to take the most nuanced of norms about candour, is understood even in the rough-and-tumble world of finance. Everyone prefers clarity to opacity, brevity to verbosity. Beyond that, one can say little about our storehouse's tools until we experiment with them.

The second contention is that moral and legal phraseology may impede ethics education. If this is not self-contradictory, it is a canard. Few of us learn that to burgle we must break and enter a domicile at night with intent to commit a felony; we all learn that we should never invade another's property. We usually do not learn morals from law books. When seeking to stimulate thinking by young

scientists, instructors select provocative discussions, not rules. What is in play is clarity. Litigants would seem no better served by norms eschewing legal distinctions than would patients by physicians' orders eschewing medical distinctions.

Nuance unbound

OSTP proposes to prohibit and define fabrication, falsification and plagiarism, and surprisingly enough to convert departure from "accepted practices" from a sufficient to a necessary condition of misconduct⁵. This seems aimed at consensus wrongs, but antixenophobia has exacted a toll in precision.

OSTP renders fabrication as "making up results and recording or reporting them". This metaphor leaves the shop sense of fabrication to differ from the regulatory only insofar as "making" differs from "making up". That puts a big load on a little adverb. OSTP goes on to define falsification as "manipulating research materials, equipment, or processes, or changing or omitting data or results such that the research is not accurately represented in the research record". Suppose that Jeeves deliberately misstates an experimental step. (A principal charge in one publicized case was failure to disclose pooling of patient samples.) According to OSTP's text, because Jeeves has not misstated or omitted either "data" or "results", no misconduct occurs. The same result follows if Jeeves deliberately misstates the value of a physical constant or signs a patent application asserting a falsehood. The cure for under-inclusiveness here does not lie in expanding the notion of "the research record", for OSTP the crux. For any academic discipline, it is mistaken to suppose that deception relates peculiarly to the record, esteemed though it be. OSTP, for example, would appropriately capture deception in oral presentations. What is fundamentally wrong about deception is harm to others, not to a medium. We need to specify the activity in the course of which harm by self-promoting deception should be forbidden. That activity we familiarly call "the conduct of research".

Rigour here is also simplicity. "Fabrication" and "falsification" are indirect ways of expressing misrepresentation. As it happens, our richest source of experiential guidance devolves from disclosure requirements of the federal securities laws⁶, thousands of judicial encounters with which have exposed many nuances of reliance on written communications. An investigator induces and betrays a listener's trust by signalling "I believe it" while believing a false utterance false or a misleading omission misleading. This understanding defines misrepresentation in the conduct of research (see box above). Here we imagine no access to private mental states, only inferences from behaviour.

OSTP would instead capture conduct committed "intentionally, or knowingly, or in reckless disregard of accepted practices". The

To capture a consensus

Research Misconduct Defined. It shall constitute research misconduct for an investigator in the conduct of research to commit misrepresentation, plagiarism, or misuse of another's work.

● The **conduct of research** consists in (i) any inquiry funded by federal funds, or (ii) the preparation or communication related to any such inquiry of (a) a document published or provided to another, including any government agency, institution of higher education, research institution, or professional organization, or (b) a formal oral presentation.

● An **investigator** is a person who participates in the conduct of research.

● A fact is **material** if it or its contrary affects the plausibility, replicability, execution, or effect on safety of a result, hypothesis, or procedure, or is such that it would be rational for others to alter their course of conduct in reliance upon it.

● **Misrepresentation** occurs if and only if an investigator, intentionally or recklessly indicating belief in the investigator's utterance, (i) delivers a false utterance concerning a material fact while

believing the utterance false, or (ii) while believing the effect misleading, omits a material fact without which the utterance is misleading.

● **Plagiarism** is the intentional presentation of the words of another as the presenter's own.

● **Misuse of another's work** is the intentional presentation as the presenter's own, without attribution appropriate for the medium, of the ideas or work of another.

council of the National Academy of Sciences has opposed offences predicated on recklessness⁷. OSTP's phrase may reflect a consensus to sanction some recklessness, but it encompasses too much. For any first author, a paper's every sentence is doubtless intentional. OSTP would exclude "honest error", but that contradicts the condemnation of recklessness and begs for a definition of "honest". That returns us to the key tests for culpability — what the investigator intentionally or recklessly indicates and what he believes.

Deceptive omissions have long been recognized as wrongful. To show that an omission renders "the research" not "accurately represented", as OSTP imagines, one must show the omission misleading. The word "material", a threshold already observed by ORI⁸, forestalls prosecution of peccadilloes. In the Imanishi-Kari case, it was held that alleged wrongdoing about the "January fusion" data concerned a "trivial" matter⁹. We may define "material" for the experimental setting (see box). OSTP would instead say that "a significant departure from accepted practices of the scientific community for maintaining the integrity of the research record" is a necessary condition for finding misconduct. This recrudescence of the "deviation" phrase is vague and — like a track umpire ranking pole-vaulters by technique instead of height vaulted — adverts to method when results control. When Amelie's paper omits the data from a botched experiment, we do not vindicate Amelie by conjuring "accepted practices" for "maintaining integrity of the research record"; we observe that the discarded data are immaterial in the context.

OSTP renders "plagiarism" as "the appropriation of another person's ideas, processes, results, or words without giving appropriate credit, including those obtained through confidential review of others' research proposals and manuscripts". Apart from doing violence to "plagiarism" (which consists only

in transcription of words), this text exonerates plagiarists who copy without appropriating and convicts non-deceptive patent infringers who appropriate without copying. OSTP would absolve copying that is not a "serious departure", but our considered understanding of plagiarism is unqualified (we do not condone "a little plagiarism"). OSTP anomalously condemns "reckless plagiarism". And its mention of work confidentially reviewed may suggest freedom to copy work otherwise reviewed¹⁰. Review ought not to be mentioned: copying is the wrong, regardless of source. We need precisely to define plagiarism and misuse of another's work (see box). If Jeeves copies by unattributed paraphrase rather than verbatim transcription, Jeeves may commit misuse.

In the norms of misrepresentation, plagiarism, and misuse of another's work, we would appear to have the simplest hypothesis explaining our considered moral judgements. In time their etymologies may seem no more alien than those of "science" and our many other borrowings from the ancients. ■

Louis M. Guenin is in the Department of Microbiology and Molecular Genetics, Harvard Medical School, 200 Longwood Ave, Boston, Massachusetts 02115, USA.

e-mail: guenin@hms.harvard.edu

Further commentary on the author's and OSTP's proposals may be found at

<http://www.hms.harvard.edu/dms/cos/guenin.html>

1. *Daubert v. Merrell Dow Pharmaceuticals*, 113 S. Ct. 2786 (1993).

2. Guenin, L. M. & Davis, B. D. *Sci. Eng. Ethics* 2, 47–54 (1996).

3. Guenin, L. M. *Acad. Med.* 71, 595–603 (1996).

4. National Academy of Sciences. *Responsible Science* (National Academy Press, Washington DC, 1992).

5. 64 Fed. Reg. 55723 (14 Oct. 1999).

6. Rule 10b-5 under the Securities Exchange Act of 1934, 17 C.F.R. § 240.10b-5.

7. Council of the National Academy of Sciences. Letter to William Raub, 15 March 1996.

8. 60 Fed. Reg. 10588 (27 Feb. 1995).

9. *In re Imanishi-Kari*, Department of Health and Human Services, Departmental Appeals Board, Research Integrity Adjudications Panel, Decision No. 1582 (1996).

10. Guenin, L. M. *Science* 271, 1790 (1996).

The Church's route to enlightenment

Our increasing understanding of the Sun's motion reflected on the Church.

The Sun in the Church: Cathedrals as Solar Observatories

by J. L. Heilbron

Harvard University Press: 1999. \$35, £21.95

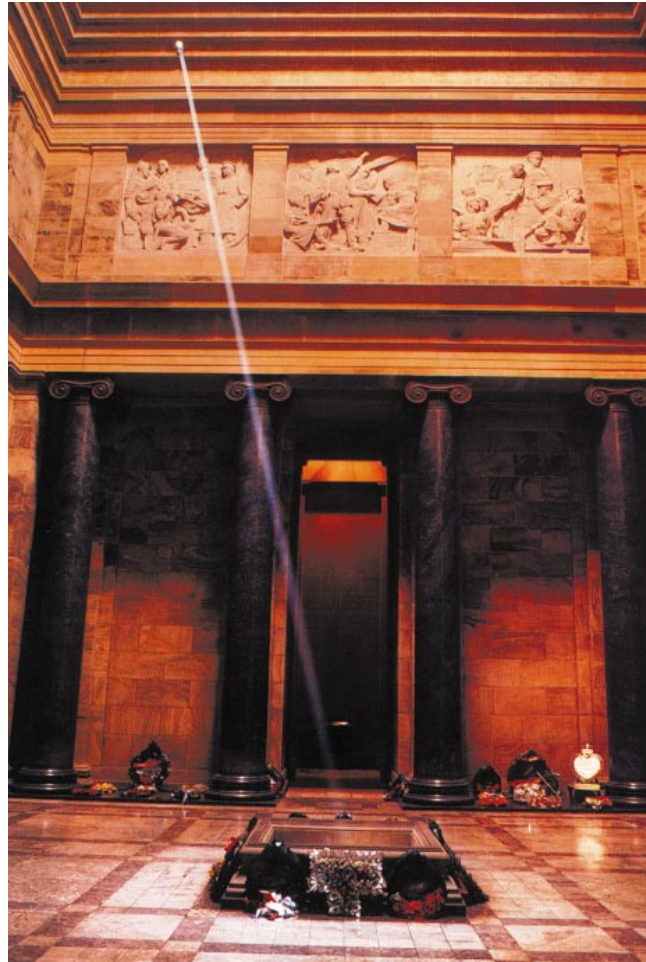
George V. Coyne

At first sight, the title of this book reminded me of similar titles by masterful storytellers, such as Willa Cather's *Death Comes for the Archbishop* and T. S. Eliot's *Murder in the Cathedral*. While it is certainly not in the traditions of fiction or fictionalized history, John Heilbron's book does tell a gripping story and with a splendid literary flair.

The story is about the early attempts to find where we are in a universe in which we think all the objects, except for ourselves, are moving. Since the search was a practical one, about how to live on the Earth, and since the Sun and the Moon are the most important objects in determining the rhythm of life on the Earth, the story is really about determining the position and motion of the Sun. But the author cleverly uses this theme to give literary unity to a much broader study in the history of science at the time when modern science was being born. By subtly inserting critical comments, the author evaluates the interactions of science in its gestation with the culture of those centuries and the repercussions that these interactions have had down to our own times. And so it becomes a story about people, and Heilbron tells it in a masterfully human way.

All his principal characters are focused upon the construction of the most effective instrument of the sixteenth and seventeenth centuries for measuring the position and motion of the Sun, the meridian. The instrument is a very simple one. It consists of a hole in the south-facing wall of an edifice which would allow a beam of sunlight to fall on a line directed north-south (the local meridian) on the floor of the edifice. Since a large, dark enclosure was needed to observe the changing position of the sunbeam on the line, churches were the obvious places to install a meridian. Churchmen encouraged and supported this because, among other practical matters, it would help in determining the calendar of Church feasts—especially Easter, which, as determined by the Gregorian calendar reform of 1582, falls on the first Sunday after the first full Moon after the vernal equinox.

Heilbron tells the fascinating story of how this was done in various churches, mainly in the Papal States, capturing with splendid irony the character of the principal personages involved. He conveys the spirit of the times and passes judgement on the larger



Point of illumination: churches were an obvious place to install a meridian.

issues that were brooding beneath the surface of the meridian. It appears that Giovanni Domenico Cassini, a closet Copernican, is Heilbron's hero or, at least, the scientist most characteristic for him of that age. He is described brilliantly: "We are led to the unappealing, unromantic, counterintuitive conclusion that Cassini led the astronomers of his time without commitment to any cosmology or world system and that he lived amicably with the persistent traditionalists of Bologna and the convinced Copernicans of Paris."

Heilbron's hallmark irony is found throughout the book. Copernicus, sent to Italy by his uncle, spent seven years "soaking up the sun and the Renaissance". His "Greek was as accurate as the Alfonsine tables". Johannes Kepler replaced the "hoary principles" of circular motion and constant velocity with "the products of exasperation, exhaustion, and genius", namely, ellipses. Cardinal Gregorio Barbarigo, who promoted mathematics and the meridian, "came within an ace of being elected pope and ended up a saint. ... The famous blue-stock-

ing of Bologna, Laura Bassi, received a licence to read Descartes and others' bad books sometime before 1740, although she suffered under the double disadvantage of being female and under age." My favourite is the picture caption: "A Jesuit telescope suspended by faith".

Those who try to interpret this book as an apologia for the Roman Catholic Church (see *The New York Times* review, 19 October 1999) are grossly mistaken. Read carefully the first two sentences of the book. With his usual brilliant irony Heilbron attributes the Church's devotion to the science of the meridian to "administration". In fact, I cannot help but think that the clever title of the book is itself an ironic judgement, and a correct one too, on the Church of those times. The Sun is our great illuminator. Did it, upon its entry through the hole of the meridian, illuminate the Church? The implied answer is obvious and, when the author discusses some of the more fundamental problems that the new science was creating for the Church, he makes it clear. He makes the correct judgement that "Urban

VIII made a bad mistake in associating himself with Galileo's condemnation and with the characterization of heliocentrism as heretical in any sense". Would that it had been a conclusion of the Galileo Commission, which instead put all of the blame on "some theologians".

In his discussion of the commission's work, Heilbron is far too lenient. For instance, in the discourses of 31 October 1992, which apparently terminated its work, the commission concluded that Galileo was wrong because he would not accept that Copernicanism was hypothetical and because he had no proofs. There is an ambiguity involved in the use of the word "hypothesis" and Heilbron is well aware of it as he alludes to it frequently, especially in his treatment of Cassini. There are two distinctly different uses of the word: a mathematical expedient to predict celestial events, or an attempt to understand the true nature of the heavens. There is no doubt that Galileo understood his own investigations to be an attempt to understand the true nature of things. It is well known that he preferred to be known as a philosopher of nature rather than as a mathematician. It cannot be denied that he sought evidence to show that Copernicanism was really true and not just a mathematical expedient. He sought to find observational verification of it and, if not with total success, with more success than had been achieved by anyone up to his day. He can certainly not be accused of "betraying the very method of which he was the inspired founder", one of the conclusions of the commission. Heilbron could have seized the opportunity to knock the commission on the knuckles a bit.

Also, he propagates a common error by claiming that Galileo's works were removed from the Index of works that Catholics should not read as a result of the Settele affair, in which a textbook on optics and Copernican astronomy went through a tortuous procedure before it received papal approval. The 1822 imprimatur on Settele's book did not refer to Galileo, nor to the sentence of 1633. It referred to the teaching of Copernicanism. And if it is claimed that the imprimatur implicitly reformed the sentence of 1633, why was that not made explicit? As a matter of fact, the works of Copernicus and Galileo remained on the Index until 1835, more than a decade after the Settele affair and even closer than Heilbron thinks to the discovery of stellar parallax, a "proof" of Copernicanism. ■

George V. Coyne is at the Vatican Observatory at Castel Gandolfo, V-00120, Vatican City, Rome, Italy.

Opening the lines of communication

Science under Socialism: East Germany in Comparative Perspective

edited by Kristie Macrakis and Dieter Hoffmann
Harvard University Press: 1999. 380 pp.
\$55, £34

Petra Werner

The history of the German Democratic Republic (DDR), which was founded on 7 October 1949 and dissolved by the unification agreement of 1990, offers the historian an enticing, discrete historical episode. A disadvantage, on the other hand, is that many of the participants are still living, making objective research more difficult. Previous analyses and attempts to integrate the history of the DDR were affected by partisan feeling and were thematically limited. The first analyses of scientific development emphasized a recent history dominated by political and social science.

The editors of *Science under Socialism*, which has appeared in English and in a modified German edition, criticize the dominance of political and social science in accounts of contemporary history, especially in previous works on the history of the DDR. First, they claim, such a one-sided view leaves out central themes in modern social development: the important role of science and its connection with the development of "productive forces" is not adequately considered. Second, it is essential to take into account the role of the scientific and technological revolution in producing propaganda.

The book is divided into four sections. In the first, on the general politics of science and research, the authors try to take an overview of the subject, with special attention to the period immediately after 1945 and to the reform attempts of the sixties. The second section covers the transformation of the Prussian Academy of Sciences into the Academy of Sciences of the DDR, as well as the role of the Leopoldina. The third focuses on the "core domains for social progress" such as biomedicine, computer science, chemistry, aircraft construction and nuclear physics.

The final section offers some interesting biographical sketches revealing the conflicts people encountered, and their lives and survival strategies as scientists. Among these, Dieter Hoffmann examines the chemist Robert Havemann, who was variously an anti-Nazi, a communist and a dissident. Mitchell Ash takes a critical look at Kurt Gottschaldt, who continued his twin research unhindered under four political systems, and asks whether this type of opportunism is typically German. Ash nevertheless confirms the high quality of Gottschaldt's science — which is an historical exception.

The book brings together authors of different ages and disciplines, of whom six are Americans and 12 Germans, nine from the former DDR. The East German participants have the complicated task of writing about a phase of their own lives. This critical insiders' view is both challenged and supplemented by the outsider's perspective of their American and West German colleagues, all the more since the editors are a former DDR citizen and an American. It is, therefore, not surprising that the preface to the English edition states, somewhat cryptically: "Writing about East German history can be controversial and complicated, especially since many of the historical actors are still alive. Whereas Kristie Macrakis wished to include a discussion of some of these problems in a prologue, Dieter Hoffmann did not; therefore we agreed that the prologue should be printed with the understanding that it is Macrakis's view and approach and not necessarily those of Hoffmann or the other contributors. Whereas Macrakis wished to discuss the creative process involved in an East-West group project and the controversy surrounding who should write East German history now, Hoffmann wanted only the finished product presented."

It is striking that the articles by the East German contributors are much less strongly judgemental. They know how difficult it is to evaluate historical processes in which one has taken part, how easy it is to be wrong and to injure individuals. The contributors from the West, on the other hand, are much less restrained in their judgements. John Connelly, an American specialist on Eastern Europe, is much less ready to admit clarifying-



The meridia at the Duomo of Palermo: the Sun at noon in spring (in Taurus) or summer (in Virgo).

ing or explanatory circumstances than the distinguished DDR historian Hubert Laitko. Connelly was born in 1960; Laitko was born in 1935 and forced to retire in 1992.

Kristie Macrakis's own two contributions reveal a predilection for the spectacular: she notes with some disappointment, in an interesting article on espionage and technology transfer, that real undercover agents rarely have the fantastic adventures that one associates with spying. She has interviewed double agent Werner Stiller (from the Ministry of State Security — the Stasi) and political celebrities such as the late Politburo member Kurt Hager and the legendary chief of the Information Division, Marcus Wolf.

The authors are diverse not only in temperament, viewpoints and expectations, but also in their approach to their subjects. Thus, a primarily sociological account of biomedical research by Rainer Hohlfeld stands next to Eckart Förtsch's theoretical systems presentation about science, higher education and technology policy, based on the interaction of partial systems and on the structure and functions of apparatus. These contrasts make the book readable and colourful.

The present volume can only provide a first step, however, towards a more powerful theoretical conception of the DDR's role in the larger system of science. Meanwhile it challenges both east and west, despite all their differences and disappointments, to continue the conversation. ■

Petra Werner is in the Alexander von Humboldt Research Group, Berlin-Brandenburgian Academy, Berlin, Germany

Hidden wealth of data in the depths

From Monsoons to Microbes: Understanding the Ocean's Role in Human Health

by the Committee on the Ocean's Role in Human Health, National Research Council

National Academy Press: 1999.

\$34.95, £24.95,

Franciscus Colijn and Sebastian Lippemeier

Many years ago, a good friend came out of the sea after a swim and looked disoriented. Had he suffered *Pfiesteria piscicida* poisoning? In fact he had lost his spectacles, although nowadays you might pick up that toxic organism when you go swimming in some US estuaries. But on the other hand, the anti-cancer agent is collected from marine organisms. These and various other aspects of humankind's relation to the sea are explored in this kaleidoscopic text, which is



A multilingual view of the sea

Deep Water (Ellipsis, £16 UK/£19 international) by Julian Caldecott and Melanie Salmon claims to be a celebration of the waterworld. The parasitic isopod shown above is just one of many photographs in a book that is published in English with French and Spanish translations on each page.

The book focuses on the need to protect the sea's fragile and inter-dependent ecosystems. It is produced by the charity Living Earth, which works through hands-on education and grassroots action to help people to resolve local environmental issues and issues concerned with human development. All royalties from the sale of the book will be used by Living Earth to fund marine-education programmes.

suited for every interested reader with a background in natural sciences.

From Monsoons to Microbes ranges broadly from the ocean's threats — natural disasters, infectious diseases and toxic algal blooms — to its possible benefits for humankind as a source of new medical and pharmaceutical products, and of marine organisms which can be used as tools for biochemical research. But the extent of humanity's interactions with the oceans is immense: no book can cover more than a fraction of it.

This book makes you aware of the potential hazards, but also of the enormous impact the ocean has on our daily life. After all, our global climate is regulated by oceanic currents and the little-understood interaction between the ocean and the atmosphere. Climatic predictions depend on our knowledge of such interactions. Recent studies have documented that we do not have sufficient insight into the regulatory mechanisms of the oceans to predict potential alterations of

ocean currents, which could lead to a new ice age within 10–50 years, an unexpected sea-level rise, a steady mean global temperature increase of 2 °C or maybe nothing at all. These uncertainties can only be overcome if interdisciplinary studies in ocean sciences continue and the cooperation of biological and physical oceanographers is reinforced.

The ocean's currents can also distribute infectious diseases to humans who eat infected shellfish, via pathogenic bacteria, viruses or toxic algal species. One wonders why these hazards do not cause more problems in coastal communities, especially in the tropics, which are by far the most vulnerable sites for health and climatic threats.

Public health problems arise most often in heavily urbanized coastal areas, where humans still tend to accumulate. This human behaviour enhances all potential hazards such as eutrophication effects, and concurrent toxic algal blooms. Without mitigation these problems may even prove to be synergistic.

New technologies within the framework of the Global Ocean Observing System (GOOS) enable us to improve our environmental observations of the oceans, which have lagged far behind those on land. In turn, this will strongly improve the predictive capabilities of physical and biological operational models.

Furthermore, we know little as yet about the extent of oceanic biodiversity. A wealth of biological information still awaits us (as long as we control pollution) in the form of marine compounds, which may prove to be of medical and pharmaceutical use. Of course, we have to start new research programmes to deal with this information, but there is a fascinating invisible world below the ocean surface. We've already reached the Moon; now it's time to take a better look inside our oceans. ■

Franciscus Colijn and Sebastian Lippemeier are at the Research and Technology Centre, Westcoast, 25761 Bissum, Germany.

Story of an obsession

Prion Biology and Diseases

edited by Stanley B. Prusiner

Cold Spring Harbor Laboratory Press: 1999.

710 pp. \$125, £???

Colin L. Masters

By my count, this is the fifth time Stanley Prusiner has edited a multi-authored book on the subject of his life-long obsession: the prion, a proteinaceous-infectious particle responsible for the transmissible spongiform encephalopathies (Creutzfeldt-Jakob disease (CJD) and kuru in humans; scrapie and bovine spongiform encephalopathy

(BSE) of domesticated herbivores). Considering that more than half of the 17 chapters are actually co-authored by Prusiner, this is the closest yet to a monograph from his laboratory. And, it has to be admitted, the book is by far the most authoritative introduction and summary of the complex data that form the basis of our understanding of the molecular basis of this esoteric group of diseases. While the human forms of the diseases are rare, their economic impact, particularly in the United Kingdom, has been profound. Moreover, their occurrence is now threatening the normal operation of blood-transfusion services worldwide. Hence the need for such an authoritative book.

The prion field has been dogged by controversy, apparently the result of the long time it takes to generate reproducible scientific data, given incubation periods that can exceed 40 years. The discovery of the prion protein (designated PrP^{Sc}) in 1982 as the principal component of the infectious particle revolutionized and galvanized investigators. Prusiner recounts the background to this, and takes us through the following critical decade in which the surrogate properties of PrP^{Sc} (insolubility, resistance to proteolysis, increased β -sheet content, polymerization into amyloid fibrils) are elucidated. A certain obfuscation remains over the issue of prion rods and amyloid filaments; since the minimal infectious unit has still not been defined, and because it appears that neither rods nor filaments are required for infectivity, this issue seems to be related to attribution of discovery.

While Prusiner and his collaborators have driven the research agenda of the molecular basis of PrP infectivity, other independent groups have complemented his efforts. The nature of infectivity has been uncovered largely through the application of transgenesis to mice, including ablation studies that prove beyond reasonable doubt that PrP plays the central role in prionogenesis. Interestingly, the recent discovery of a duplicated homologue of PrP may necessitate some re-interpretation of these knockout studies, although leaving the central hypothesis intact. Studies in yeast by Reed Wickner and colleagues show that accessory chaperones (possibly the postulated 'protein X' of mammalian prions) may facilitate the conversion of a normal host-cell protein into a pathogen. Many enigmas remain: the topology of PrP in the cell membrane, the mechanisms underlying nerve-cell degeneration, the partial creation of *de novo* infectivity in overexpressing mutant PrP in mice, and the failure yet to create infectivity *in vitro*.

The emergence of BSE and then the threat of an epidemic in humans caused by contamination of the food chain has assured the future of intensive research of these diseases. The book closes with several helpful practi-

cal chapters on methodologies (including insights into future therapeutic strategies), antibody probes (which may prove essential for future clinical tests) and biosafety issues. The latter area is now of immediate concern, since the general levels of ignorance about the spread of CJD and BSE cause regulatory authorities to respond by containing the worst-case scenario, often with drastic financial implications.

Overall, the volume is virtually a monograph by Prusiner, and as such presents a

very unified picture of the current state of research. It is the best comprehensive overview of this field and is an ideal introduction for anyone wishing to launch a career in prionology. Perhaps the major message would be that Prusiner achieved his successes through outstanding collaborators, most of whom are represented in the book. ■

Colin L. Masters is in the Department of Pathology, University of Melbourne, Parkville, Victoria 3052, Australia.

Science in culture

The Brain

A theatre show from Forkbeard Fantasy which will be touring Britain from 6 December and throughout much of next year

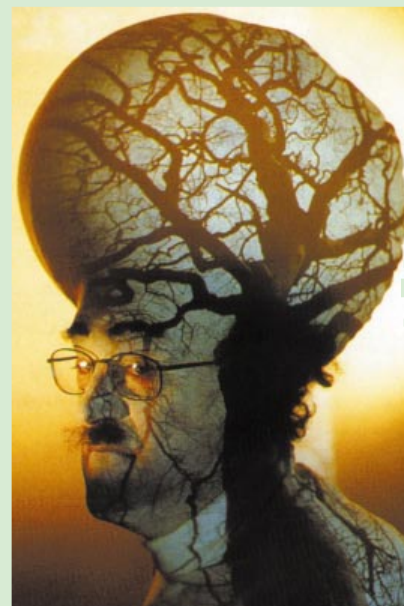
Bonnie Hurren & Robert Meech

Making sense of *The Brain*, an experimental performance from Forkbeard Fantasy, is like trying to make sense of the brain itself. Ronald C. James's well-known photograph of a Dalmatian dog against a dappled background reminds us just how easily the brain finds order in a random world. In the world of Forkbeard Fantasy, on the other hand, anarchy is created effortlessly out of a high degree of order. The technical skill required to mount this mixed-media performance is immense. Video machines, film projectors, neon lights, music, an 'intelligent' lighting system, automated smoke machines (to create a sensation of smell) all work seamlessly to give the illusion of chaos.

The Brain was funded in part by an award from the Wellcome Trust's Science on Stage and Screen competition. The nine winners of this, the first year of the competition, each received up to £36,000 (US\$56,500) to develop a production illuminating some aspect of medical science. *The Brain* was written and directed by Paul B. Davies, who has had considerable success on the Edinburgh Fringe — the offbeat offshoot of Britain's major arts festival — as well as on radio.

"Logical thought has no vantage point from which to observe its own workings," according to Davies in his programme notes, "but the imagination can propel the mind outside itself. ... *The Brain* is one such acrobatic enterprise and although we can't pretend that it contains much real science, ... modern scientific knowledge about the brain is the springboard from which we have launched it."

The one cerebral moment in all the confusion is when Emil Toescu of Birmingham University (billed as neuroscience adviser to Forkbeard Fantasy) appears on screen as a talking head to explain the neurophysiological basis of the sense of smell. He ends with a good joke, one of many good jokes during this 90-minute production. Look out for the one about that Dalmatian dog.



But for all the chaos, there are themes. Smell is important in the narrative, as are bees and phantom limbs. And behind all the fantasy there are important points about common oversimplifications made in the name of making science entertaining. Forkbeard use ridicule as a powerful weapon: how do you cure someone with a phantom limb? Answer: cut his head off.

Much intelligence and wit have gone into the production. Most of the ideas are well integrated but a few seem shoehorned into place, which makes for a disjointed narrative. But then narrative is the last thing on Forkbeard's mind.

Forkbeard acknowledge works by Richard Gregory, Rita Carter, Susan Greenfield, V. S. Ramachandran, Sandra Blakeslee, Marcel Proust and the children's comic *The Beezer*. ■
Bonnie Hurren is a freelance actor and director, c/o Bristol Old Vic Theatre School, 2 Downside Road, Clifton, Bristol BS8 2XF, UK. Robert Meech is in the Department of Physiology, University of Bristol, University Walk, Bristol BS8 1TD, UK. *The Brain* will have its final venue at the Natural History Museum, London, between 16 and 30 September 2000. For performance details write to Forkbeard, PO Box 1241, Bristol BS99 2TG, UK. It is suitable for sophisticates seeking to nourish their inner child.

The forgotten menace

Nuclear weapons stockpiles still represent the biggest threat to civilization.

Paul Doty

The passage from one millennium to the next is a powerful stimulus to reflect on our most vital problems. Top of the list must be the legacy that this century bequeaths to the next and to the millennium beyond — the risk that the tens of thousands of nuclear weapons left over from the Cold War will bring an end to civilization.

While many informed people felt this threat during the Cold War, a sense of relief from imminent danger has been the hallmark of the first post-Cold War decade. As the concern over a global apocalypse has subsided it has been replaced by the threat of the use of one or a few weapons by accident, by terrorists or by 'rogue' nations. This refocusing is understandable: it follows from the natural human concern with the immediate and the difficulty in dealing with events that are unlikely but much more catastrophic.

Lost in this shift of focus is the tremendous change of scale that it brings about. The use of a few weapons could mean the destruction of a few cities. This is alarming against the background of no similar violence for decades. Clearly its prevention deserves attention. But the loss would be far below that suffered in the Second World War. Despite the anguish in affected regions, the physical damage could be repaired, the fabric of civilization would remain. How strikingly different would be the consequences of war among nuclear powers escalating to the use of most of their stockpiles of weapons.

For comparison, the damage caused in the Second World War is comparable to that of roughly 100 nuclear weapons if used to maximum effect. What, then, if the destruction were hundreds of times greater? Nuclear devastation would be delivered within a brief time rather than cushioned over several years, leaving great regions permanently contaminated with radioactivity, the world's electric circuitry destroyed, along with most life-sustaining infrastructure, and the schools, hospitals, libraries, museums and monuments of the world's cultures obliterated.

Some people would survive initially. How many in the long term would depend on the extent of radioactive contamination, the ability to reorganize life in isolated, less affected regions and the will of the survivors to start over.

The many major wars of this millennium offer little hope that the next one will not provide the triggers that would lead to large-scale nuclear war if today's weaponry remains or is replaced. To prevent this outcome — and avoid losing our civilization in the next millennium (or century) — two deep and radical changes must occur.

First, the world's stockpile of nuclear weapons must be rolled back to the level of 100, thereby limiting the damage, should they be used, to an amount from which recovery is possible. Second, the nuclear nations, and those that seek such status, must come to see how irresponsible the risk of excessive nuclear weapons is, and appreci-

ate how such excesses diminish, rather than enhance, their military security.

Why aim for 100 weapons, and not zero? The reason lies in the enormity of the problems that this would bring into play, diluting the effort to reach the 100 level. The consensus needed to reach a nuclear-free world may remain politically unattainable. To attain it would require an inspection system that would be enormously complex, intrusive and expensive. And it would be fundamentally unstable, because of its vulnerability to hidden or quickly assembled weapons.

The reduction of nuclear weaponry is difficult and expensive. But few people know that nearly half of the world's deployed nuclear weapons have been dismantled in the past decade. To go beyond dismantlement, and ensure that the removed fissile material cannot be recycled into new weapons, involves complex disposal processes that are now beginning. The dismantling of nearly 20,000 nuclear weapons might suggest that the process could simply be continued. But that cannot be the case, for several reasons.

Dismantling so far has largely been undertaken by Russia and the United States because, with the end of the Cold War, their arsenals were seen as excessive by any standards and because it was more costly to maintain them than to eliminate them. Importantly, the reduction left so many active weapons that it was not necessary to verify the numbers. Further reductions will require reliable verification and an assessment of the remaining inventory. Moreover, the nuclear powers other than Russia and the United States will have to join the effort. This presents a very ambitious agenda in nuclear arms control.

But reducing numbers only gives an index of whether politics in the nuclear nations is evolving in a way that justifies decreasing reliance on nuclear weapons as an instrument of policy. There is little evidence that policy in Russia and the United States is evolving in this direction. Therefore, the threat of keeping our civilization a hostage to an irresponsible and outmoded view of the role of nuclear weapons must be met by an informed public. Although now somnolent, it is here that the awakening must occur. To quote from the preface to the Unesco charter: "Wars begin in the minds of men and it is there that the defences of peace must be built."

Paul Doty is at the John F. Kennedy School of Government, Harvard University, Cambridge, Massachusetts 02138, USA.



Fingers off the button? Leaders of nuclear powers, such as US President Bill Clinton and Chinese President Jiang Zemin, must realize that their weapons diminish, rather than enhance, their security.

Spawn of Satan?

Widespread egg donation has led to a generation of superkids.

Nicola Griffith

Egg donation has begun to bias the new and controversial Raswani Social Intelligence scale and the more traditional Stanford-Binet IQ test. Nationwide testing has been prompted by what some educators are calling an exponential change in the behaviour of kindergarten children. "We noticed a real big difference," said Anita Cummings, a teachers' assistant at a grade school in an upscale Chicago neighbourhood. "When I first took this job in 2015 we'd sometimes get a kid who was not only smart, but smart all around, who really knew how to handle other kids. Now we have half a dozen or more every year. They're almost perfect. To tell you the truth, they're a little frightening."

A local bus driver, who prefers to remain anonymous, was less circumspect. "Spawn of Satan," he said. "It ain't natural. These kids climb on the bus, say 'please' and 'thank you', and read all the way to school. Lord knows, I can't abide all the yellin' and runnin'-up-and-down of normal kids but this ain't natural."

Demographics point to white, middle-class women in the 45–48 age range. "So it's only to be expected," said a harassed-looking Dr Judith Sternberg, returning from testimony to a congressional ethics sub-committee. "Record numbers of career-oriented, well-educated women are now choosing, in their mid-forties and older, to have children. And they're choosing extremely smart, well-educated women in their twenties to be the donors. The rest is genetics."

"Nonsense," responds educational sociologist Mike Chattergee, "the deciding factor is the child's upbringing. These older mothers tend to be more affluent, so they can give the infant everything it needs in the way of education and nurture. Better nurturing means a happier, healthier, more well-adjusted child."

A corollary of the egg-donation boom is the change in behaviour noted in the spouses of married mothers. "It's good old-fashioned competition," said red-faced Jack Donatelli at the bowling alley in Midwich, Connecticut. "Women get to pick the father so, if you can't hack it, move aside for someone who can."

"There's nothing old-fashioned about it," says his companion, who would only give his name as Bill. "Look at me, see that muscle? Strong as an ox. Good job. But that's not enough any more. Now it's 'Oh,



Bill, why don't you do the laundry while I write that proposal for the UK office?" Because I don't want to, that's why, but do I say anything? No. Because if I do she'll get someone else to father her goddamn test-tube baby, and I'll be slaving to support a cuckoo child!"

Some Church leaders have long decried the commercialization of egg donation. "Life is a gift from God," said the Archbishop of Chicago. "When one woman who is blessed with fertility can help bring joy to another, that gift should be given freely." Unitarians, on the other hand, believe God is in everything, even the test-tube and the bank account. Other religions, such as Islam, forbid the procedure entirely.

But ethics have nothing to do with it, argues the director for reproductive health at the Women's Clinic. "It's all very well wishing things were different," says Dr Allison Toomin, "but this is the real world. What healthy 22-year-old college senior in her right mind would go through a month of pain, daily injections, bloating, hormonal disturbance and the risk of medical complications, just for altruism? Especially when prospective mothers are offering \$75,000

and a trip around the world if the donor has a SAT score of over 1,500, good looks and perfect health."

Senator George W. Bush III believes people like Dr Toomin are wrong. "It's just not right," he tells voters gathered at a rally in Texas. "These women are *buying* smarts for their babies. They're *buying* fertility, love and a secure old age, just because they were too selfish to stop working and have babies when they were young and healthy, while God-fearing folks are so crippled by Big Government taxes that they can hardly scrape together the bread for their own little ones!"

Not far from where Bush is speaking lies Austin, one of the epicentres of the intelligence spike now being observed all over the country. Others include Seattle, the San Francisco Bay Area, and certain neighbourhoods of larger cities such as Atlanta and Boston. "With the exception of Atlanta, these are all very white cities," points out M'Shelle N'dele Mbele, from the Urban Justice Center in St Louis. She grins sardonically: "Wonder why that is."

She believes that, like most racial issues, this is at heart a money-based discrimination. Few would disagree: reproduction by egg donation is expensive but, with a first-time success rate now approaching 80 per cent, it's by far the most reliable of *in vitro* technologies.

Dr Sternberg believes that, as Americans approach the second half of the twenty-first century, egg donation is here to stay. "What I told the ethics sub-committee is that we need to think about what this means for business. Global competition from emerging nations is threatening the ascendancy of American corporations. We need our female executives to remain focused on their jobs through their thirties and early forties and not be distracted by the idea of a biological clock. Egg donation lets us reset that clock, if not banish it all together. Right now America has the edge. Egg donation lets us keep it."

The sub-committee has declared that there will no longer be regulation of viable human ova at the federal level, and the debate is under way regarding tax credits for donors and clinics.

All parties expect controversy. For every egg donor and prospective mother there will be someone like old-timer Sam Underhill, overheard recently at the Green Dragon Inn in Bywater, Maine. "It's not natural," he said, "and trouble will come of it." ■

Nicola Griffith (www.sff.net/people/Nicola) has just finished writing *Red Raw*, the sequel to her latest novel, *The Blue Place*. She lives in Seattle, Washington.

OLIVER BURSTON

Giving a boost to atoms

Kristian Helmerson

Matter waves can now be amplified in the same way that a laser amplifies light. Such matter-wave amplifiers will be essential for future developments in atom optics.

It is difficult to imagine modern life without amplifiers. Whether your stereo is faithfully reproducing a Mozart concerto, or your cell phone is boosting a wireless call, amplifiers are used widely in daily life. Amplifiers, as the name suggests, amplify or enhance a signal. Typically, the signal is a wave, such as an oscillating electrical current, a radio wave or a beam of light. Quantum mechanically, matter can also be regarded as waves. Hence it seems reasonable to imagine that, like other waves, matter waves can be amplified. Such direct amplification of atomic waves has now been observed by a group at MIT. On page 641 of this issue, Inouye *et al.*¹ report phase-coherent matter-wave amplification using a Bose–Einstein condensate.

According to the quantum or wave-mechanics picture of matter, atoms have a wavelength that depends inversely on their speed. For atoms at room temperature, this wavelength is less than a nanometre. In addition, the atoms zip around in all directions and the probability that two atoms are in the same place, at the same time, with the same velocity, is very small. This makes it difficult to observe the wave properties of atoms in everyday experience. A Bose–Einstein condensate is a quantum gas in which many of the particles share the same quantum state²; that is, they are in the same place, at the same time, with the same velocity. So a Bose–Einstein condensate behaves as a macroscopic wave. At the ultracold temperatures (tens of nanokelvins) characteristic of a Bose–Einstein condensate, the atoms' wavelengths are tens of micrometres and can be observed with conventional optics. Such a large number of atoms having the same long wavelength makes a Bose–Einstein condensate the ideal system in which to demonstrate matter-wave amplification.

The basic function of an amplifier is to take a wave (the input) and then produce a more intense, but otherwise identical, wave (the output). In practice, an amplifier has a gain medium where there is a transfer of external energy (the pump source) to the output according to the input. The amount by which the output increases compared to the input is the gain of the amplifier. In an optical amplifier, or laser, the gain medium is a collection of atoms or molecules (gas, liquid or solid) with energy levels such that a transition between them is at the wavelength of light to be amplified. The pump source

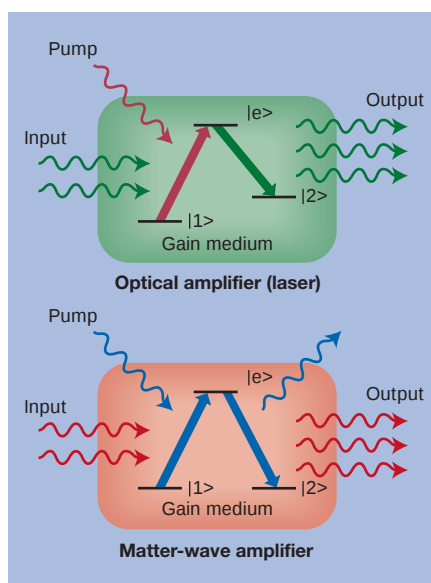


Figure 1 Optical waves or matter waves? In an optical amplifier or laser, the gain medium consists of atoms or molecules with at least three energy levels. Atoms in $|1\rangle$ are transferred to $|e\rangle$ by a pump source, which could be another laser. The input wave stimulates the atoms to transfer to $|2\rangle$ and emit a photon. The emitted photon is indistinguishable from the input photons. So a more intense, but otherwise identical, optical wave emerges from the output. In the matter-wave amplifier, three distinct energy levels or states of the atoms are required. Unlike light, the atoms in the matter-wave amplifier cannot simply be created, so the gain medium also serves as a reservoir of atoms. After absorbing a 'pump' photon, an atom in excited state $|e\rangle$ can emit a photon and end up in a different state (such as $|2\rangle$ with a random phase). But in the presence of the input matter wave the excited atom is more likely to end up in $|2\rangle$ with the same velocity and phase as the input atoms. So a more intense, but otherwise identical, matter wave emerges from the output. Amplification will end when all the atoms in the gain medium have been transferred to $|2\rangle$.

transfers some of the atoms into the higher of the two energy levels. The input wave then stimulates this population to transfer to the lower energy level, emitting radiation in the same direction and at the same wavelength as the input wave (Fig. 1). This process is called stimulated emission, and explains the name laser: light amplification by stimulated emission of radiation.

The matter-wave amplifier works on the same principle of stimulated emission, but with stimulated emission of atoms, not radiation. Unlike light, atoms cannot be created out of a vacuum, so you need a reservoir of atoms to be emitted. Inouye *et al.*¹ prepared their input matter wave using optically induced Bragg diffraction³ of a Bose–Einstein condensate. This is a process whereby atoms are exposed to two laser beams with slightly different frequencies. The atoms, initially at rest, absorb a photon from the higher frequency beam and emit into the lower frequency beam. The result is that the atoms get a momentum kick equal to the difference in momentum of the two beams, which is given by their wavelength and relative directions. Hence an input matter wave is produced which recoils in a chosen direction.

Amplification of the input wave is achieved by 'pumping' the Bose–Einstein condensate with an intense pulse from one of the lasers used for Bragg diffraction; here the condensate is behaving as the gain medium (Fig. 1). Some of the condensate atoms absorb a photon from the laser beam, and if they 'see' the input matter wave, before they can radiate the absorbed photon by spontaneous emission, they will be induced to emit photons along the direction of the other laser beam. As a result, the recoiling atoms will have the same momentum (the same direction and wavelength) as the input wave. In other words, the input wave has been amplified by stimulated emission, producing a larger output wave from the pumped gain medium. Inouye *et al.* observe typical gains of 10 to 100, which means there are 10–100 times more atoms in the input pulse after amplification.

An amplifier that can produce an output wave that is more intense than, but otherwise identical to, the input wave is called phase coherent. This distinguishes such amplifiers from those that simply increase the strength of the output. This might be acceptable if you simply wanted a high-power light beam to cut material. But phase-coherent amplification is important in applications that are phase sensitive, such as interferometry or phase-modulated signal transmission. Inouye *et al.* used interferometry to check the phase coherence of their amplifier. They used optically induced Bragg diffraction to produce a reference-

matter pulse identical to the input pulse. By looking at the interference of this reference pulse with the amplified pulse, the researchers verified that the process was phase coherent.

Previous studies have observed stimulated emission without seeing phase coherence. The ability of optically pumped condensate atoms to serve as a gain medium for matter-wave amplification was first observed by the MIT group⁴. They showed that a pumped condensate, initially radiating photons by spontaneous emission, had sufficient gain to amplify emission in a random direction. This process is analogous to the amplified spontaneous emission of light in atomic vapours. Stimulated emission of atoms is also demonstrated by four-wave mixing of matter waves⁵. In this case the stimulated emission of atoms is mediated by the mean-field interaction of the atoms in a Bose–Einstein condensate, and not by pumping from an optical field.

The demonstration by Inouye *et al.*¹ that matter waves can be amplified and maintain their phase is an exciting development in the rapidly growing field of matter-wave optics. Already, there are reports of similar experiments from researchers at the University of Tokyo (M. Kozuma, personal communication). But the task is not easy. Inouye *et al.* estimate a gain of four for phase-coherent amplification (compared to a number gain of 10–100). Atoms, unlike photons, can interact with each other directly, leading to distortion of the amplified matter wave. Nonetheless, such experiments are perhaps the closest matter-wave equivalent of the optical laser. The first generation of ‘atom lasers’⁶ was based on the coherent extraction of a beam of atoms from a Bose–Einstein condensate. But some argued that the analogy with optical lasers is incomplete because you cannot really amplify atoms. The

demonstration of phase-coherent matter-wave amplification by stimulated emission of atoms is much closer to a true atom laser.

It is difficult to imagine the future impact of such studies — few would have predicted the success of the modern laser. The earliest optical lasers were clunky devices that languished in labs for years before anyone found a use for them. But a coherent matter wave is inherently more difficult to work with than a beam of light. For example, there is no material equivalent to glass that a beam of atoms can readily pass through, and the atoms need to be maintained in an ultrahigh vacuum. But there may be times when matter waves are better than laser light. For instance, phase-coherent optical amplifiers are used in optical interferometer gyroscopes to measure gravity accurately. Matter-wave interferometer gyroscopes are already more sensitive than their optical counterparts⁷, but could be further improved by incorporating a matter-wave amplifier. Phase-coherent optical amplifiers are also used to boost the output of a laser, as in fibre-optic communications. Similarly, a phase-coherent matter-wave amplifier could be used to boost the output of an atom laser, for example in atom lithography. Such amplification of matter waves may be the future of better and brighter atom lasers. ■

Kristian Helmerson is at the Atomic Physics Division, National Institute of Standards and Technology, 100 Bureau Drive, Stop 8424, Gaithersburg, Maryland 20899-8424, USA.
e-mail: kristian.helmerson@nist.gov

1. Inouye, S. *et al.* *Nature* **402**, 641–644 (1999).
2. Townsend, C., Ketterle, W. & Stringari, S. *Phys. World* 29–34 March (1999).
3. Kozuma, M. *et al.* *Phys. Rev. Lett.* **82**, 871–875 (1999).
4. Inouye, S. *et al.* *Science* **285**, 571–574 (1999).
5. Deng, L. *et al.* *Nature* **398**, 218–220 (1999).
6. Helmerson, K., Hutchinson, D., Burnett, K. & Phillips, W. D. *Phys. World* 31–35 August (1999).
7. Gustavson, T. L., Bouyer, P. & Kasevich, M. A. *Phys. Rev. Lett.* **78**, 2046–2049 (1997).

Alzheimer's disease

The tangled tale of tau

E. Mandelkow

Alzheimer's disease is the main form of dementia in today's ageing population. It is characterized by abnormal protein deposits in the brain, such as the extracellular amyloid plaques and the intracellular neurofibrillary tangles. The tangles are made up of a protein called tau, which is modified by phosphorylation and aggregated into so-called paired helical filaments. Because hyperphosphorylation seems to occur before aggregation, it is considered one of the earliest signs of neuronal degeneration. But which protein kinases are responsible for it?

Two papers in this issue go some way

towards answering this question. Tsai and colleagues¹ (page 615) report that the cyclin-dependent kinase 5 (Cdk5) is deregulated in the brains of people with Alzheimer's disease. Not only that, but Cdk5 also hyperphosphorylates tau and accumulates in tangle-bearing neurons. And Greengard and colleagues² (page 669) show that Cdk5 can alter signalling processes in neurons by altering the balance between its kinase and phosphatase activities.

The tau protein is associated with neuronal microtubules. It stabilizes them and regulates the transport of vesicles or

organelles along them, supports the outgrowth of axons and serves as an anchor for enzymes³. The human central nervous system contains six isoforms of tau, generated from a single *tau* gene by alternative splicing. Embryos contain only one (the smallest) isoform; the larger isoforms are induced only when neurons differentiate and cell processes (axons and dendrites) are generated. Tau is a basic protein, rich in serine and threonine residues, and it can be phosphorylated by many kinases. A common hypothesis is that phosphorylation weakens the affinity of tau for microtubules. As a result tau detaches and then aggregates, eventually leading to neuronal degeneration (Fig. 1).

Which kinase initiates this chain of events? Proteins related to the cell-cycle kinase Cdc2 came under suspicion early on for several reasons. First, tau from the brains of patients with Alzheimer's disease is phosphorylated at serine or threonine residues followed by a proline. This points to the activity of a ‘proline-directed’ kinase, which includes members of the Cdc2 family, but also other kinases such as the mitogen-activated protein kinases or GSK-3. Second, one Cdc2-family kinase, Cdk5 (also known as NCLK, for neuronal Cdc2-like kinase), is one of two main tau-protein kinases isolated from brain tissue (the other being GSK-3b⁴). Third, proline-directed phosphorylation of tau increases strongly during mitosis, and correlates with other markers of mitosis⁵. This is consistent with the idea that degeneration could be the response of a differentiated neuron to inappropriate mitotic signals, which forces that neuron to commit suicide (apoptosis). Indeed, inhibitors of Cdc2-like kinases can prevent an apoptotic response⁶.

Arguments against this circumstantial evidence include the fact that proline-directed phosphorylation of tau does not strongly influence its functions (binding to microtubules, for example) compared with the effects of other kinases such as the micro-

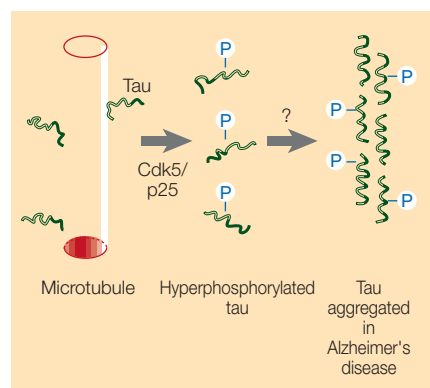


Figure 1 Hyperphosphorylation and aggregation of tau. Phosphorylation of tau is thought to weaken its affinity for the microtubules to which it normally binds. Tau then detaches and forms the intracellular aggregates characteristic of Alzheimer's disease.

tubule-affinity regulating kinase or protein kinase A⁷. Moreover, cells transfected with Cdk5 show no real change in tau phosphorylation. Nonetheless, Tsai and colleagues¹ now provide compelling evidence for the involvement of Cdk5 in the phosphorylation of tau.

The authors first show that Cdk5 is not properly regulated in the brains of people with Alzheimer's disease. Normally the kinase is activated by binding a partner protein, p35 (also known as Nck5a). The function of p35 is analogous to that of the cyclins, which bind to other Cdc2-like kinases. The activity of Cdk5 is tightly regulated through spatial and temporal constraints (Fig. 2) — p35 is usually attached to the plasma membrane via a myristoylation anchor at its amino terminus. It has a short lifetime and is degraded rapidly via the ubiquitination–proteasome pathway.

Surprisingly, all of these constraints are lifted when p35 is cleaved into a smaller, amino-terminal part (p10) and a larger, carboxy-terminal part (p25), although the mechanism of this aberrant cleavage is not yet known. The p10 portion disappears rapidly, but p25 remains bound to the kinase. This p25–Cdk5 complex is constitutively active, long-lived and no longer restricted to the plasma membrane. In other words, Cdk5 can now go out on a 'phosphorylation spree', phosphorylating everything in sight — including tau. In support of this, Tsai and colleagues have observed tau co-precipitated with p25 and Cdk5 in the degenerating neurons of Alzheimer's disease.

Other substrates of Cdk5, such as the proline-rich domains of neurofilaments⁸, might suffer a similar fate. Indeed, deposits of highly phosphorylated neurofilaments have been found, together with Cdk5, in patients with Parkinson's disease and amyotrophic lateral sclerosis⁹. The hypothesis linking Cdk5 to neurofibrillary deposits is attractive for another reason — both Cdk5 and tau are normally involved in the development of axons and dendrites. The level of Cdk5 increases transiently during brain development, then decays again. However, levels of Cdk5 remain high in regions with the highest plasticity, such as the entorhinal region or the olfactory system¹⁰. This correlates with the regions in which the hyperphosphorylation and aggregation of tau are first seen in Alzheimer's disease¹¹.

Many questions remain. For example, what is the physiological function of Cdk5? Is it regulation of the neuronal actin cytoskeleton, as suggested by Tsai and colleagues' results¹? Or does Cdk5 modulate signalling cascades via kinase/phosphatase activities, as indicated by Greengard and colleagues' study²? These authors showed that Cdk5 turns a modulator protein called DARPP-32 (dopamine and cyclic AMP-regulated phosphoprotein, relative molecular mass 32,000) into an inhibitor of protein kinase A. Another

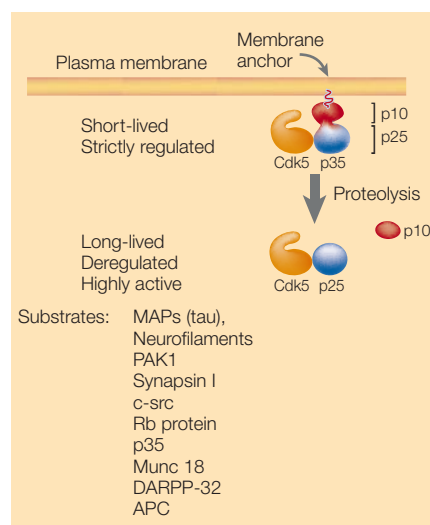


Figure 2 Regulation of the cyclin-dependent kinase 5 (Cdk5). Cdk5 is activated by binding to p35, which anchors it in the plasma membrane. If p35 is cleaved, however, Tsai and colleagues¹ have found that the Cdk5–p25 complex is a highly active, long-lived form of the enzyme. It can break free from the plasma membrane and phosphorylate a number of substrates, including tau and, as Greengard and colleagues² have shown, DARPP-32. (MAPs, microtubule-associated proteins; Rb, retinoblastoma; APC, adenomatous polyposis coli gene product.)

study¹² implies that Cdk5 may control axonal transport. But all of them bear on the observed influence of Cdk5 on neuronal differentiation¹⁰.

Which protease is responsible for the aberrant cleavage of p35? Does the hyperphosphorylation of tau by Cdk5 really cause microtubule breakdown and tau aggregation, or are these indirect effects involving other kinases? After all, microtubule dynam-

ics in axons is determined not only by tau, but by other microtubule-associated proteins as well, and phosphorylation of tau does not necessarily lead to aggregation¹³. How does the activity of Cdk5 relate to the newly discovered¹⁴ tau mutations that cause fronto-temporal dementias without changing the phosphorylatable residues? What is the crosstalk between tau phosphorylation by Cdk5 at a serine–proline motif and proline isomerization¹⁵?

Answers will come with time but, independently of them, Tsai and colleagues' results provide potential targets for diagnosis or therapy. These may come either at the level of p35 cleavage or at the level of Cdk5 inhibition, and therapy should be possible with the development of highly specific inhibitors.

E. Mandelkow is at the Max-Planck-Unit for Structural Molecular Biology, Notkestrasse 85, D22607 Hamburg, Germany.
e-mail: mandel@mpasmb.desy.de.

- Patrick, G. *et al.* *Nature* **402**, 615–622 (1999).
- Bibb, J. *et al.* *Nature* **402**, 669–671 (1999).
- Mandelkow, E.-M. & Mandelkow, E. *Trends Cell Biol.* **8**, 425–427 (1998).
- Imahori, K. & Uchida, T. *J. Biochem. (Tokyo)* **121**, 179–188 (1997).
- Vincent, I., Jicha, G., Rosado, M. & Dickson, D. W. *J. Neurosci.* **17**, 3588–3598 (1997).
- Padmanabhan, J., Park, D., Greene, L. & Shelanski, M. *J. Neurosci.* **19**, 8747–8756 (1999).
- Drewes, G., Ebner, A. & Mandelkow, E.-M. *Trends Biochem. Sci.* **23**, 307–311 (1998).
- Sharma, P. *et al.* *J. Biol. Chem.* **274**, 9600–9606 (1999).
- Nakamura, S. *et al.* *Neurology* **48**, 267–270 (1997).
- Delalle, I., Bhide, P., Caviness, V. & Tsai, L.-H. *J. Neurocytol.* **26**, 283–296 (1997).
- Braak, H. & Braak, E. *Acta Neuropathol.* **82**, 239–259 (1991).
- Ratner, N., Bloom, G. & Brady, S. T. *J. Neurosci.* **18**, 7717–7726 (1998).
- Schneider, A., Biernat, J., von Bergen, M., Mandelkow, E. & Mandelkow, E.-M. *Biochemistry* **38**, 3549–3558 (1999).
- Wilhelmsen, K. *Proc. Natl Acad. Sci. USA* **96**, 7120–7121 (1999).
- Lu, P.-J., Wulf, G., Zhou, X. Z., Davies, P. & Lu, K. P. *Nature* **399**, 784–788 (1999).

Volcanology

Magma fragmentation in eruptions

Dork Sahagian

Some volcanic eruptions are gentle and effusive, and even (as on Hawaii) support a tourist trade. Others, such as the eruptions of Mount St Helens in 1980 and of Mount Pinatubo in 1991, are violent and terrifying, causing severe damage, loss of life and even temporary climate change. The difference between them stems from the nature of the erupting magma and the way that it breaks up into fine particles and ash by a process called fragmentation.

Papers by Zhang¹ and by Marti *et al.*² (pages 648 and 650 of this issue) provide new insights into magma behaviour during fragmentation. Zhang describes a new criterion to assess brittle failure of magma. Any brittle

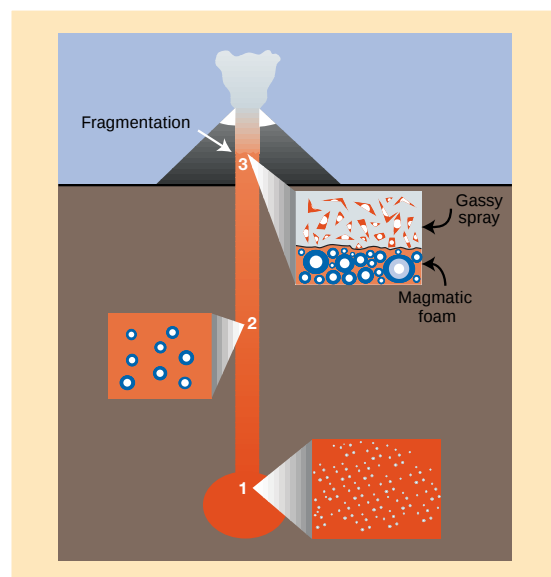
material can support stress (force) by undergoing strain (bending or stretching) up to its elastic limit, after which it either bends permanently, or breaks by brittle failure. When the thin walls between bubbles in a magmatic foam are stressed sufficiently during bubble growth, they break by brittle failure, and the bubbles coalesce. Marti *et al.* delve into the details of the stress state within the magma before and during fragmentation by examining the morphology of pumice from a volcano in the Andes.

It is bubbles generated deep within a volcano that are the driving force behind eruptions. They grow by diffusion of dissolved magmatic gas and decompression on the way

Box 1: Birth and death of bubbles

The bubbles that drive a volcanic eruption form (1 in the figure here) by the poorly understood process of nucleation⁹. Indeed, in principle, they should not exist at all: surface tension should cause the gas pressure inside a very small bubble to be so much greater than that of the surrounding magma that the gas should dissolve into the magma and the bubble should disappear. It is not known what causes enough gas (water vapour and carbon dioxide) to diffuse 'upstream' against a concentration gradient in a pure magma to allow a bubble to form despite surface tension, but it may involve chemical structures within the magma that affect surface tension locally.

Once bubbles exist, their growth is controlled by inward diffusion of gases^{10,11}, as well as decompressive expansion as the magma rises up a chamber or conduit system (2). The pressure-dependent solubility is reduced and the gas moves down a concentration gradient from a high point midway between two bubbles, and into the bubbles. The rate of diffusion is characteristic of the magma and depends on the same properties as the



viscosity (more viscous magmas have slower diffusivities).

If a magma rises to the surface fast enough, the diffusion of dissolved gas through the magma between the bubbles may not proceed quickly enough to maintain chemical equilibrium with the falling ambient pressure. So the magma may become supersaturated in gas at the low pressure of the volcano's vent or eruptive column (less than 1 atmosphere), and bubbles may grow swiftly at that point. The bigger they and

their surface area become, the faster the rate of gas influx. Fuelled by a sufficiently high level of magma supersaturation, this runaway feedback effect can lead to explosive bubble growth. At some critical growth rate⁷, the magma no longer behaves as a fluid, but as an elastic solid — as Zhang¹ and Marti *et al.*² describe. When this magmatic foam can no longer support the stresses induced by bubble growth, as at 3, brittle failure of the bubble walls causes fragmentation into a gassy spray.

D. S.

up to the Earth's surface^{3–6}, as described in Box 1. But fluid magmas (even highly viscous ones) should allow bubbles to grow until their internal pressure matches that of the exterior (1 atmosphere) before fragmentation, and thus lead to more gentle fragmentation and less explosive eruptions than are observed.

This apparent discrepancy is reconciled by the theoretical formulation and laboratory observations of Zhang¹ and of Marti *et al.*². They show that as the viscous material deforms rapidly due to bubble growth, it can no longer be treated as a fluid with a Newtonian viscosity, but assumes the characteristics of a brittle material. In part, brittleness may be exacerbated by cooling of the bubble walls by the latent heat of exsolution of water evaporating from the magma, so that they may even turn to glass in extreme cases⁷. With brittle behaviour of magmatic foams, fragmentation does not depend so much on the total amount of strain (as for typical solids),

but rather on the strain rate.

The key to brittle fragmentation during eruptions is for the bubbles to grow and deform so quickly that there is no time for the magma to 'relax' as would a fluid. It thus behaves like a brittle solid that can be broken, rather than bent or stretched. Although one normally considers the total amount of strain (bending or stretching) as a failure criterion for elastic solids, magmas do not behave quite like solids. If a small stress is applied, fluid flow will eventually relieve that stress (relax) as the fluid flows (strains) to accommodate it. But when the stress is large enough, the strain rate to accommodate it is too large for the magma, and it breaks like an elastic solid.

Zhang finds that, under sufficiently high strain rates, a magma can be treated as an elastic solid, and he applies brittle-failure theory to an idealized case of spherical bubbles surrounded by spherical shells of magma. As the bubbles grow and cause thinning of the walls of magma that separate

them, the tensile stresses grow in the magma until they surpass its tensile strength, causing fragmentation. As further evidence for the brittle behaviour of magma under high strain rates, Marti *et al.* find that the morphology of tube pumices (volcanic foams with highly elongated bubbles) records the state of strain at the point of fragmentation. With this strain marker, they demonstrate that the magmatic foam from which the pumice was derived deformed as a fluid until shortly before fragmentation, then sheared in a visco-elastic manner just before fragmenting as an elastic material. They conclude that fragmentation was associated with a great increase in strain rate.

The disruption of a magmatic foam⁸ marks a fundamental change in the topology of a two-phase magmatic system as the system shifts from a bubbly melt to a gassy spray. A magma containing bubbles is a continuous 'liquid' melt with isolated gas bubbles that can communicate only through diffusion of gas through the melt. After fragmentation, however, the system becomes a continuous gas that is carrying tiny droplets or fragments of magma (which do not communicate with each other at all). Further, because of the low viscosity of the gas, the droplets may have a different velocity from that of the gas, and so may be segregated. In some cases, this can lead to *nuées ardantes*, devastating, high-density 'avalanches' of gas and ash particles that can move downslope very quickly and over long distances — as, for instance, one did on Mount Pelée, Guadeloupe, destroying the coastal town of St Pierre and its inhabitants.

With this new understanding of fragmentation, volcanologists can reconcile the violent behaviour of many volcanoes with the characteristics of the erupted magmatic materials, leading us another step closer to predicting the nature of eruptions. As described in Box 1, however, the question of bubble nucleation remains. Once bubbles exist, we know how they grow. Once they grow, we now know how they 'pop'. But we still don't know how they form in the first place.

Dork Sahagian is at the Institute for the Study of Earth, Oceans and Space, Morse Hall, University of New Hampshire, 39 College Road, Durham, New Hampshire 03824, USA.

e-mail: dork.sahagian@unh.edu

1. Zhang, Y. *Nature* **402**, 648–650 (1999).
2. Marti, J., Soriano, C. & Dingwell, D. B. *Nature* **402**, 650–653 (1999).
3. Alidibirov, M. & Dingwell, D. B. *Nature* **380**, 146–148 (1996).
4. Alidibirov, M. & Panov, V. *Bull. Volcanol.* **59**, 481–489 (1998).
5. Gardner, J. E. *et al. Bull. Volcanol.* **58**, 144–162 (1996).
6. Papale, P. *Nature* **397**, 425–428 (1999).
7. Sahagian, D. L. & Proussevitch, A. A. *J. Volcanol. Geotherm. Res.* **74**, 19–38 (1996).
8. Proussevitch, A., Sahagian, D. & Kutolin, V. J. *Volcanol. Geotherm. Res.* **59**, 161–178 (1993).
9. Toramaru, A. *J. Geophys. Res.* **100**, 1913–1931 (1995).
10. Proussevitch, A. & Sahagian, D. *J. Geophys. Res.* **103**, 18223–18251 (1998).
11. Proussevitch, A. A. & Sahagian, D. L. *J. Geophys. Res.* **101**, 17447–17456 (1996).



100 YEARS AGO

The current number of the *Lancet* contains an article of particular interest dealing with the effects upon the wounded of the Mark II., the Mauser, the Dum-dum, and the Mark IV. bullets. The article, which has been written by Dr. Arthur Keith and Mr. Hugh Rigby, gives a clear idea of the relative amount of destruction caused by each of these modern military bullets, and the experiments upon which the writers' views are founded, confirm fully the experiences which have already been reported from the seat of war in South Africa. A glance at the illustrations shows the terrible havoc wrought by the Mark IV. and Dum-dum bullets and shows also that the old Martini-Henry bullet made an enormous and jagged wound compared with the neat little track that is left behind the Mark II., which our forces are using in South Africa, or the Mauser which is being used by the Boers. Dr. Keith and Mr. Rigby have not, however, been able to obtain results in their experiments with Dum-dum bullets that endorse Prof. Von Bruns's statement of the case against the English open-nosed bullet. All open-nosed bullets cause fearful injuries, but it is contended that Prof. von Bruns must have used Dum-dum bullets of an exceptional nature to get the results which he recorded.

From *Nature* 7 December 1899.

50 YEARS AGO

Samuel Johnson once observed how quietly people will endure an evil which they might at any time very easily remedy; he gave as an instance a family which had possessed an island for more than four hundred years and never made a commodious landing-place, though a few men with pick-axes might have cut a staircase out of the rock in a week's time. At the recent meeting of the British Association at Newcastle upon Tyne, Section J (Psychology) held a discussion on the "Human Problems in the Design of Machinery and Working Methods", and it was emphasized that, although knowledge is increasing, even modern communities have as yet little scientific understanding of these questions. In fact, Dr. Johnson's remark still holds good to-day, because many instances were quoted at Newcastle in which ordinary common sense has been lacking in the planning of the physical, psychological and social environments in which men work.

From *Nature* 10 December 1949.

Ecology

Water fleas on cycles

James P. Grover

Abundances of animals vary over time, and explaining these fluctuations is a central goal of ecology. Especially intriguing are the many cases where populations cycle¹. The period of cycles is rarely one year, ruling out obvious tracking of the seasons. This leaves at least two possibilities to explore. Predators may over-exploit their prey, leading to cycles of population explosion and collapse, with prey numbers rising to high levels during the latter phase. Alternatively, partially synchronized reproduction might drive demographic cycles with alternating large and small cohorts of offspring, leading to cycles in adult numbers. On page 653 of this issue², McCauley *et al.* tease apart these possibilities for an organism with well-known population cycles — the common water flea *Daphnia*.

In thinking about population cycles, ecologists often turn to mathematics³. In mathematical models, predator-prey cycles generally have high amplitudes⁴, which are greatly exaggerated by enriching the prey's habitat⁵. In contrast, theory predicts that cycles driven by demography and phased reproduction lack the high amplitude and sensitivity to enrichment characteristic of predator-prey cycles.

In field studies, it is difficult to disentangle the mechanisms driving population cycles from the many other confounding sources of natural variation, and experimental control is often infeasible. Model systems of predator and prey cultured in the laboratory have been popular — such as *Daphnia* and the microalgal prey that it filters from the water. The properties of this system would be expected to give classical predator-prey cycles. *Daphnia* populations can double within weeks and over-exploit their prey, clearing natural waters of algae⁶. Algae reproduce even more rapidly, doubling within days, and easily reach high abundance when *Daphnia* are sparse. But decades of study of *Daphnia* and algae have not shown clear evidence for predator-prey cycles.

Early laboratory studies of *Daphnia* populations did indeed reveal cycles⁷, but regular additions of algal food prevented over-exploitation. Instead, *Daphnia* reproduction became synchronized. When sparse, adults obtained a large ration of food and produced numerous offspring. Adults of the subsequent generation were abundant, obtained a much smaller ration, and left fewer offspring. Cycles with relatively low amplitude, and out-of-phase variations of juveniles and adults, were the result. Extending these studies, McCauley and his colleagues² have



Figure 1 Reproduction in *Daphnia*. Normal eggs are produced parthenogenetically and incubated in the dorsal brood chamber (top). Resting eggs are provisioned with energy stores and enclosed in a protective brood case (bottom). McCauley *et al.*² found that these resting eggs have the effect of damping the species' population cycles.

conducted experiments and combed the literature on natural populations of *Daphnia*. Their work supports the presence of demographic, but not predator-prey cycles. Why is the basic biology of this predator and its prey, which strongly hints of over-exploitation and prey escape, so misleading? Previous painstaking work by this group has eliminated several hypotheses⁸.

One remaining hypothesis postulates that algal species relatively invulnerable to ingestion by *Daphnia* have a selective advantage in rich habitats. Inedible algae compete for nutrients with edible algae, and effectively reduce the richness of the habitat for edible algae, dampening the tendency to cycle. McCauley and colleagues achieved the difficult task of preventing inedible algae from proliferating in their laboratory systems. Doing so induced classical predator-prey cycles, and this result does much to settle the question of why *Daphnia* do not show such cycles under less controlled conditions. This finding probably generalizes to many predator-prey systems: invulnerable competitors of prey should be highly favoured by natural selection, and therefore common in nature.

ARDEA

Perhaps more intriguing, when inedible algae were absent, only some of the experimental cultures displayed predator–prey cycles of high amplitude. Others cycled with lower amplitude, and with phasing of adults and juveniles suggesting demographic cycles. This result demonstrates ‘alternative attractors’ — that is, two dynamical states that characterize the long-term behaviour of *Daphnia* populations, with subtle differences in initial conditions determining which state is reached. The nonlinear mathematical models used to study population cycles often predict such phenomena, but most ecologists have probably despaired of verifying their existence in real populations.

McCauley and colleagues also reveal a biological mechanism governing the switch between the two types of population cycles. *Daphnia* are parthenogenetic — they reproduce without sex — and their normal eggs quickly hatch into juveniles. *Daphnia* also produce, by sexual reproduction, resting eggs for long-term survival (Fig. 1). This diverts energy from normal reproduction and could prevent population explosions and over-exploitation, thus turning off the predator–prey cycle. McCauley and colleagues tested this possibility by replacing mothers carrying resting eggs one-for-one with mothers carrying normal eggs. This nearly doubled the amplitude of population cycles.

Production of resting eggs essentially disperses offspring to the future, and is analogous to dispersing them in space. In theory, spatial dispersal could stabilize or destabilize population dynamics, and the range of possibilities for temporal dispersal requires further study. Long-term population dynamics could depend on the hatching of resting eggs. Indeed, others have drawn attention to similarities between the ‘egg banks’ of aquatic animals and the ‘seed banks’ of plants⁹. Dispersal to the future is a widespread feature among animals, in the form of resting eggs, diapausing stages (in which animals go into a kind of suspended animation), and long-lived adults. It could stretch out the timescale of ecological dynamics, blurring the distinction with evolutionary dynamics. The study by McCauley *et al.* should do much to stimulate research on its implications in many types of organisms. ■

James P. Grover is in the Department of Biology, University of Texas, Arlington, Texas 76019, USA. e-mail: grover@uta.edu

- Kendall, B. E., Prendergast, J. & Bjornstad, O. N. *Ecol. Lett.* **1**, 160–164 (1998).
- McCauley, E., Nisbet, R. M., Murdoch, W. W., de Roos, A. M. & Gurney, W. S. C. *Nature* **402**, 653–656 (1999).
- Volterra, V. *Nature* **118**, 558–560 (1926).
- May, R. M. *Science* **177**, 900–902 (1972).
- Rosenzweig, M. L. *Science* **171**, 385–387 (1971).
- Lampert, W., Fleckner, W., Rai, H. & Taylor, B. E. *Limnol. Oceanogr.* **31**, 478–490 (1986).
- Slobodkin, L. B. *Ecol. Monogr.* **24**, 69–88 (1954).
- Murdoch, W. W. *et al. Ecology* **79**, 1339–1356 (1998).
- Hairton, N. G. Jr *Limnol. Oceanogr.* **41**, 1087–1092 (1996).

Oceanography

Climate and the Gulf Stream

Jean-Claude Duplessy

The Gulf Stream is one of the most enduring features of the North Atlantic circulation, and was well known to ships’ masters sailing between Europe and America. On their return voyage to Europe, they sailed with this warm current through the Straits of Florida. They followed it to about 35° N and then headed for Europe along with the westerly winds (Fig. 1). On page 644 of this issue¹, Lynch-Stieglitz *et al.* show that during the Last Glacial Maximum, some 21,000 years ago, the speed of the Gulf Stream was much lower than it is today. Cold climates, therefore, seem to coincide with the reduced transport of warm water to the high latitudes of the North Atlantic.

The Gulf Stream consists of warm water arriving from the equatorial and tropical regions through the Gulf of Mexico. It is driven by the Trade Winds, which cause sea levels to be higher in the western part of the Atlantic basin and generate an area of high water pressure centred on the Sargasso Sea, comparable to the anticyclonic cells of the atmosphere. Just like winds, water tends to flow from areas of high pressure to areas of low pressure. However, the Coriolis force, caused by the Earth’s rotation, deflects the flow to the right in the Northern Hemisphere. Ocean currents, therefore, describe a clockwise gyre in the tropical Atlantic, and as the higher sea level is found in the west, the current that flows along the American coast is narrow and fast.

The current is particularly intense and deep in the Florida Straits, through which water leaving the Gulf of Mexico is forced to flow. The narrowest part of the straits is

between Florida in the west and the Bimini Island in the east, where it forms a 80-km-wide trench with maximum depth of 800 m. The speed of the current can reach up to 1.5 m s⁻¹ in the strait, and the Coriolis force tends to push the water towards Bimini. This force creates a local pressure and density gradient, which results in a strong tilt in the surfaces of constant temperature and density (see Fig. 2 on page 645). The faster the current, the stronger the tilt.

Lynch-Stieglitz *et al.*¹ use the oxygen-isotope composition of bottom-dwelling foraminifera in a set of cores collected on both sides of the Florida Straits to show that the tilt of the surfaces of constant temperature or density was significantly less during glacial conditions than it is today. Both the ¹⁸O content of foraminiferan shells and seawater density increase as a result of increasing salinity and decreasing temperature, and they are linked by an empirical relationship. Lynch-Stieglitz *et al.* established a similar relationship for the glacial ocean, taking into account the salinity and ¹⁸O increase due to build-up of continental ice-sheets. Their isotope measurements indicate only a small temperature and density contrast in the Florida Straits during the Last Glacial Maximum, and they therefore infer that the Gulf Stream was 35% weaker then.

The observation suggests that at that time ocean circulation was sluggish. The Gulf Stream is part of the global ocean circulation (the conveyor belt), which in the Atlantic today carries warm surface water northwards and cold, dense deep water southwards. The net result of the conveyor belt’s



Figure 1 Chart of the Gulf Stream drawn by Timothy Folger in the second half of the eighteenth century. The map was made on the instructions of Benjamin Franklin, Postmaster-General of the Colonies, for the benefit of ships carrying mail between Britain and America.

operation is the transfer of heat, not only from equatorial regions to high northern latitudes, but also from the Southern to the Northern Hemisphere. A huge amount of heat is released to the atmosphere as the conveyor belt's surface water cools and sinks to the abyss, and it keeps Europe much warmer than Canada at the same latitude.

During the Last Glacial Maximum, the Earth's climate was much colder than today and the strongest cooling (of 10 °C or more) was in the North Atlantic and over Europe^{2,3}. Geochemical and sedimentological studies of deep-sea sediments show that the production of North Atlantic Deep Water, the deep southward-flowing limb of the conveyor, was strongly reduced during glacial conditions⁴⁻⁶. Computer models have reproduced the main features of the glacial Atlantic circulation and have shown that the annual mean meridional heat transport was reduced by about 30% at low latitudes and close to zero north of 45° N, allowing the cold polar winds to blow unabated over Europe⁷. Lynch-Stieglitz *et al.* provide the first evidence that the low-latitude surface branch of the conveyor belt was less active during glacial conditions, and that the weaker Gulf Stream resulted in a large reduction in both northwards and interhemispheric transport of heat.

Obviously, however, determination of past oceanic currents by the tilt of surface of constant density is limited by certain assumptions. First, such reconstructions assume that wind patterns were not significantly different in the past. Atmospheric models driven with glacial boundary conditions support this hypothesis^{8,9}, but they do not apply to the remote past, when the geography was different from today. Second, reliable estimates of water transport require large variations in density, which are found

only in warm upper waters. The density contrast is much smaller in the cold, deep ocean and cannot be determined accurately given the experimental errors in the measurements of oxygen isotope ratios in foraminifera.

Nevertheless, the method pioneered by Lynch-Stieglitz *et al.* shows great promise for the reconstruction of past patterns of oceanic water transport, in particular during the sudden fluctuations in temperature that occurred every 1,000 years or so during the last glacial period. There seems to be little doubt that these abrupt changes are due to massive iceberg discharges and changes in the Atlantic's conveyor^{10,11}, although the associated changes in ocean circulation and northward heat transport are still poorly understood. Nevertheless, there is no doubt that understanding the behaviour of the Gulf Stream during these events will shed new light on the mechanisms of climate change, which may well act in the near future. ■

Jean-Claude Duplessy is at the *Laboratoire des Sciences du Climat et de l'Environnement, Laboratoire mixte CNRS-CEA, Parc du CNRS, 91198 Gif sur Yvette cedex, France.*
e-mail: Jean-Claude.Duplessy@lsce.cnrs-gif.fr

1. Lynch-Stieglitz, J., Curry, W. B. & Slowey, N. *Nature* **402**, 644-648 (1999).
2. CLIMAP Geol. Soc. Am. Map and Chart Ser. MC-36 (Geol. Soc. Am., Washington DC, 1981).
3. Guiot, J., Pons, A., de Beaulieu, J. L. & Reille, M. *Nature* **338**, 309-313 (1989).
4. Duplessy, J.-C. *et al. Paleoceanography* **3**, 343-360 (1988).
5. McCave, I. N., Manighetti, B. & Beveridge, N. A. S. *Nature* **374**, 149-152 (1995).
6. Sarnthein, M. *et al. Paleoceanography* **9**, 209-267 (1994).
7. Fichefet, T., Hovine, S. & Duplessy, J.-C. *Nature* **372**, 252-255 (1994).
8. Lautenschlager, M. & Herterich, K. *J. Geophys. Res.* **95**, 22547-22557 (1990).
9. Kutzbach, J. *et al. Quat. Sci. Rev.* **17**, 473-506 (1998).
10. Oppo, D. W. & Lehman, S. J. *Science* **259**, 1148-1151 (1993).
11. Vidal, L. *et al. Earth Planet. Sci. Lett.* **146**, 13-26 (1997).

Cell biology

Glutamate primes the pump

Patrik Rorsman and Erik Renström

Glutamate is the most widely distributed excitatory neurotransmitter in the central nervous system. It is stored within nerve terminals, in small vesicles that are released when the nerve cell is stimulated. Once released, glutamate diffuses to neighbouring nerve cells where it binds to glutamate receptors, resulting in an influx of sodium and calcium ions (ionotropic receptors) or mobilization of intracellular calcium (metabotropic receptors). On page 685 of this issue, Maechler and Wollheim¹ report that, as well as having an extracellular function in neurotransmission, glutamate acts as an intracellular signal. They show that in

insulin-secreting pancreatic β -cells, glutamate is involved in preparing the insulin-filled secretory granules for release — a process commonly referred to as 'priming'.

Glucose stimulates the pancreatic β -cells to secrete insulin after an increase in the intracellular concentration of calcium. We have known for over 30 years² that glucose stimulation elicits a biphasic secretory response consisting of an initial transient phase and a second, sustained component (Fig. 1). The most common form of human diabetes is associated with abnormalities in this release pattern, so we need to establish the underlying cellular mechanisms.

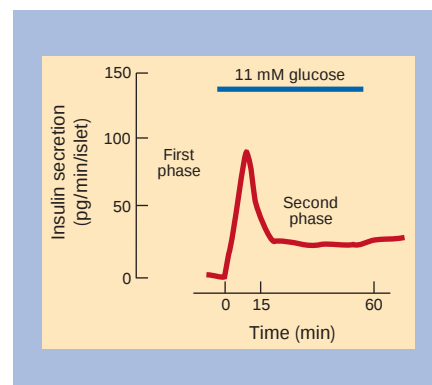


Figure 1 Biphasic stimulation of insulin secretion in response to glucose. The response consists of an initial transient phase in which the readily releasable pool of secretory granules is depleted, followed by a second, slower phase in which this pool is replenished by granule 'priming'.

The two phases of insulin secretion are thought to reflect the sequential release of two functional subsets of secretory granules³. In this model, the rapid release and depletion of a readily releasable pool of granules (RRP), containing only about 50 of the cell's 13,000 or so secretory granules, accounts for the first phase. The second, slower phase reflects the time-dependent replenishment of this pool by priming of granules originally in the reserve pool. Priming proceeds at a rate of 5-10 granules per minute per β -cell, and probably involves a product of cell metabolism because this phase is evoked only by stimuli that can be metabolized (such as glucose), and not by agents that merely increase the levels of intracellular calcium³. It has not been possible to establish the identity of this factor, but Maechler and Wollheim¹ now propose that glutamate, generated by amination of the Krebs-cycle intermediate α -ketoglutarate, is the elusive metabolic signal.

How might glutamate promote granule priming? The release of neurotransmitters and many hormones (such as insulin) involves fusion of synaptic vesicles or secretory granules, respectively, with the plasma membrane (exocytosis). This requires an interaction between various fusion proteins collectively referred to as SNAREs ('SNAP receptors', where SNAP stands for soluble NSF-attachment protein)^{4,5}. According to the SNARE model, exocytosis is preceded by ATP-dependent priming of secretory granules by the *N*-ethylmaleimide-sensitive fusion protein (NSF)⁴. Yet although insulin secretion depends on ATP, perturbation of NSF activity has little effect on release⁶.

This discrepancy can now be explained by Maechler and Wollheim's findings, which indicate that ATP hydrolysis by a granular proton pump, rather than by NSF, is more prominent in granule priming. The electrochemical proton gradient that results from

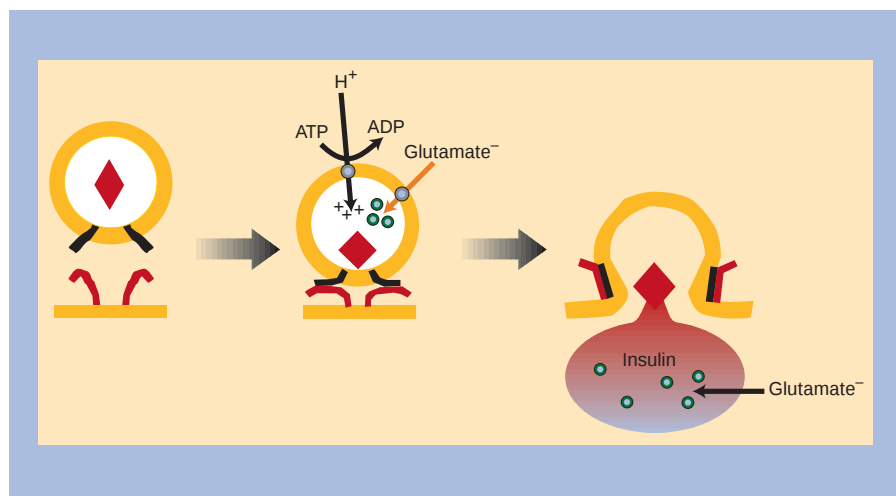


Figure 2 Model to account for the effects of glutamate on granule priming, based, in part, on the results of Maechler and Wollheim¹. The granular membrane contains v-SNAREs (black), whereas the plasma membrane contains t-SNAREs (red). The two membranes come together, and granule priming requires intragranular acidification. Glutamate uptake reduces the granular membrane potential, facilitating the pumping of H⁺ ions into the granule.

the activity of this proton pump can be envisaged as driving negatively charged glutamate into the secretory granules (although we still do not know how uptake of glutamate facilitates release). The identity of the glutamate-uptake mechanism in the β -cell is also not known, but pharmacological studies indicate that it is related to the mechanism documented in synaptic vesicles⁷. Maechler and Wollheim propose that glutamate acts by reducing the granular membrane potential, in which case it would act as a negative counter-ion, allowing a larger pH gradient to develop across the granular membrane (Fig. 2). In the absence of such a counter-ion, the large granular membrane potential that developed on proton pumping (positive inside) would quickly prevent further uptake of H⁺ and, thus, acidification.

It seems wasteful of the β -cell to use glutamate in a process where it could easily be replaced by chloride or any other cytosolic anion. One possibility is that, once priming has been completed, glutamate also has an extracellular function similar to that in the central nervous system. Interestingly, ionotropic glutamate receptors have been documented in all types of endocrine cell in the pancreatic islet⁸. This points to the exciting possibility that glutamate coordinates the complex interplay between the different types of islet cell and their secretory products.

With the paper by Maechler and Wollheim, researchers should finally be able to get a firm grip on biphasic insulin secretion and its metabolic regulation. Secretion involves similar mechanisms in different cell types. So the functions that the authors have documented in the β -cell — namely, the role of glutamate and a low intragranular pH in the priming of secretory granules — may have counterparts in other secretory cells.

Bafilomycin (an inhibitor of the vesicular H⁺-ATPase) has, in fact, been reported⁹ to abolish a late phase of glutamate release in neurons. In this context, it may be significant that vacuole acidification is required for the

pairing of the SNARE proteins in yeast¹⁰. If this also applies to the SNAREs involved in the fusion between the granules and plasma membrane, then it is easy to see how granule acidification promotes secretion. The traditional view that a low intragranular pH is required for hormone processing and uptake may represent only one side of the coin. And as the coin is now tossed, we may discover that intragranular acidification is also essential for exocytosis itself. ■

Patrik Rorsman and Erik Renström are in the Department of Molecular and Cellular Physiology, Lund University, Sölvegatan 19, S-223 62 Lund, Sweden.

e-mail: patrik.rorsman@mphy.lu.se

1. Maechler, P. & Wollheim, C. B. *Nature* **402**, 685–689 (1999).
2. Curry, D. L., Bennett, L. L. & Grodsky, G. M. *Endocrinology* **83**, 572–584 (1968).
3. Eliasson, L. *et al.* *J. Physiol.* **503**, 399–412 (1997).
4. Hanson, P. L., Heuser, J. E. & Jahn, R. *Curr. Opin. Neurobiol.* **7**, 310–315 (1997).
5. Lang, J. *Eur. J. Biochem.* **259**, 3–17 (1999).
6. Kiraly-Borri, C. *et al.* *Biochem. J.* **314**, 199–203 (1996).
7. Özkam, E. D. & Ueda, T. *Jpn. J. Pharmacol.* **77**, 1–10 (1998).
8. Weaver, C. D., Yao, T. L., Powers, A. C. & Verdoorn, T. A. *J. Biol. Chem.* **271**, 12977–12984 (1996).
9. Zoccarato, F., Cavallini, L. & Alexandre, A. *J. Neurochem.* **72**, 625–633 (1999).
10. Ungerman, C., Wickner, W. & Xu, Z. *Proc. Natl Acad. Sci. USA* **96**, 11194–11199 (1999).

Nuclear physics

Taking serious risks seriously

Sheldon L. Glashow and Richard Wilson

Risk management is not a new idea: miners once used caged canaries as methane detectors. But modern technologies have made worldwide catastrophes imaginable — disasters so dreadful that we demand proof-in-principle that they cannot happen. Failures at nuclear reactors and chemical plants (such as Union Carbide in Bhopal, India) can kill or endanger many thousands of people. So the old protocol for risk avoidance — try it once; if it turns out to be dangerous don't do it again — is no longer acceptable. For example, scientists working on the Manhattan Project seriously considered whether a nuclear explosion could release enough energy to ignite the Earth's atmosphere. Their theories said no, and history has proved them right. But do we know enough about genetic engineering to proceed safely, or could someone unwittingly (or even deliberately) create a plague worse than the Black Death? For now, it's the physicists who are under the spotlight. The worry is that a new particle accelerator could trigger an irreversible process that would destroy our planet. It is a fair concern: one that must be raised, and one that has been answered decisively by scientists in the United States¹ and in Europe².

This latest doomsday vision relates to a unique facility now being completed at Brookhaven National Laboratory (Fig. 1). At the Relativistic Heavy Ion Collider (RHIC), beams of highly charged gold or lead atoms (the heavy ions) travelling at relativistic speeds (99.95% of light speed) will speed in opposite directions around circular race-tracks before colliding. RHIC is truly an atom smasher: nucleus–nucleus impacts, taking place thousands of times per second, will each produce thousands of secondary particles. These incredibly complex 'events' will be recorded by sophisticated detectors and analysed at supercomputers or farmed out to a world consortium of smaller computers. RHIC will study matter at densities and temperatures never seen in the laboratory. On a small scale, it will reproduce the extreme conditions that reigned in the early Universe, conditions under which the constituents of ordinary matter are expected to be liberated as a quark–gluon plasma. Physicists have long speculated about this state of matter, but RHIC may soon let them glimpse it.

Meanwhile, the media and a concerned public demand to know whether these experiments could have unforeseen adverse consequences. The root of their anxiety may



Figure 1 Would RHIC wreck the world? Fears that the Relativistic Heavy Ion Collider at Brookhaven in the United States could trigger an unforeseen disaster have been allayed by new calculations^{1,2} that show the risk of a worldwide catastrophe to be truly negligible.

have been a comment by theorist Frank Wilczek in the July issue of *Scientific American*³, which was picked up by a British newspaper. Later that month, the director of Brookhaven got together a panel of independent experts (including Wilczek) to investigate the reality behind the headlines. The report from Buzna *et al.*¹ identifies three conceivable disaster scenarios at RHIC: in which experiments produce 'black holes' that could gradually consume the Earth; a 'vacuum instability' that could expand catastrophically in all directions at the speed of light; or 'strangelets' — a stable kind of 'strange matter' — that would grow to incorporate ordinary matter, perhaps transforming the entire Earth into its form.

The first two issues have been raised, and dismissed, each time a new particle accelerator opens. Using similar arguments, Buzna *et al.* were able to conclude that neither poses any threat at RHIC. There is no chance at all that RHIC could manufacture a black hole or a gravitational singularity. Even if the RHIC (or its higher-energy successors) could create a black hole, it would be so tiny that it would evaporate instantly. Previous studies also argue against a vacuum instability, but cannot quite rule it out. In the natural world, relativistic heavy ions in the form of cosmic rays have been making RHIC-like collisions with one another in space for aeons (more, in fact, than will ever take place at RHIC). These distant collisions do not make RHIC experiments any less useful because they cannot be directly studied, but one fact is clear: cosmic-ray collisions in space have not led to the creation of a new vacuum, so we can breathe easily.

The third possibility is a new concern raised by the fact that RHIC accelerates

heavy ions rather than individual elementary particles, and must be considered more carefully. This was done by Buzna *et al.*¹ and also by Dar *et al.*² at CERN in Geneva. Both groups include theorists who were among the first to speculate that lumps of strange matter called strangelets — which contain many strange quarks as well as the usual up and down quarks that make up atomic nuclei — might be more stable than ordinary matter. If strangelets exist (which is conceivable), and if they form reasonably stable lumps (which is unlikely), and if they are negatively charged (although the theory strongly favours positive charges), and if tiny strangelets can be created at RHIC (which is exceedingly unlikely), then there just might be a problem. A newborn strangelet could engulf atomic nuclei, growing relentlessly and ultimately consuming the Earth. The word 'unlikely', however many times it is repeated, just isn't enough to assuage our fears of this total disaster.

Once again, Mother Nature's own experiments with energetic cosmic rays have much to teach us. These natural processes produce collisions similar to those to be studied at RHIC, and in much greater numbers. Using different, but decisive, cosmic-ray arguments, the two groups^{1,2} make a compelling case that RHIC will not produce dangerous strangelets. Buzna *et al.* use the Moon as their canary. Lacking a protective atmosphere, with a surface rich in mid-sized atoms such as iron, it is a plausible target on which incident cosmic rays of iron (or larger) nuclei with RHIC energies could produce strangelets. Yet countless collisions over billions of years have left the Moon intact. In contrast, Dar *et al.* consider heavy-ion collisions in space that are virtually identical to those at RHIC. Strangelets produced in this way would be swept up into stars where they could instigate supernova explosions or create ultra-bright stars (such as have never been seen). The low rate of supernovae — a few per millennium per galaxy — also makes this extremely unlikely.

Both of these groups, using worst-case arguments and sound empirical data, find the chances of a catastrophe at RHIC to be truly negligible: "Cosmic ray collisions provide ample reassurance that we are safe from a strangelet-initiated catastrophe at RHIC"¹, and "Beyond reasonable doubt, heavy-ion experiments at RHIC will not endanger our planet"². Amen. Even though the risks were always minimal, it is reassuring to know that someone has bothered to calculate them. Now, the only conceivable disaster at RHIC would be a costly failure to detect the fabled quark–gluon plasma. ■

Sheldon L. Glashow and Richard Wilson are at the Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138, USA.

e-mails: glashow@physics.harvard.edu
wilson@physics.harvard.edu

1. Buzna, W., Jaffe, R. L., Sandweiss, J. & Wilczek, F. <http://xxx.lanl.gov/hep-ph/9910333>
2. Dar, A., De Rujula, A. & Heinz, U. *Phys. Lett. B* (in the press). <http://xxx.lanl.gov/hep-ph/9910471>
3. Wilczek, F. *Sci. Am.* **281**, 8 (1999).

Neurobiology

The topography of memory

Howard Eichenbaum

For more than 40 years neuroscientists have explored the cerebral cortex with microelectrodes, recording the electrical activity of single neurons while 'tickling' them with different stimuli. Early investigators found that the cortical areas responsible for initial processing of sensory information yielded their secrets relatively easily^{1,2} — for each area, a small set of simple sensory fea-

tures that activate the neurons could be identified. Many of these features have since been found to have a systematic anatomical organization; they are mapped onto the surface of the cortex, and the topographies for different sensory features are interleaved within each cortical area.

Since those early successes, the feature-coding properties of other cortical areas,

including late stages of sensory³ and cognitive⁴ processing, have also been characterized. Yet there has been limited success in describing a topography for those cortical areas⁵, suggesting that higher and more complex aspects of perception and cognition might not be systematically mapped. On page 610 of this issue, however, Hampson and colleagues⁶ reveal a functional organization for the hippocampus, one of the highest cortical-processing areas in the brain.

Although the hippocampus is crucial for memory, the nature of its contribution is controversial. Human studies point to critical hippocampal function in 'declarative memory'⁷ (our ability to record personal experiences and weave these episodic memories into our knowledge of the world about us). By contrast, animal studies indicate that the hippocampus is dedicated to 'spatial mapping'. Animals with hippocampal damage show severe and selective impairments in spatial memory⁸ (their ability to learn and

express knowledge of spatial relations among cues in the environment). Animal studies have also revealed the existence of hippocampal 'place cells' — neurons that fire when the animal is at a particular location, as though signalling occupancy of a coordinate in a spatial map⁹.

Recent studies offer the beginnings of a reconciliation between the 'declarative' and 'spatial' views of hippocampal processing. For example, rats with hippocampal damage have impaired learning, and they cannot express knowledge about relations between distinct experiences in nonspatial memory¹⁰. It has also become clear that hippocampal neurons are activated in association with specific nonspatial cues, as well as with cognitive events and behaviours that occur at different times in many places¹¹. These observations suggest the hippocampus is fundamental in registering and linking episodic memories into larger networks for both spatial and nonspatial memory¹². In a

network of spatial memories, the critical links may be remembered places that a person visited at different times. In a nonspatial memory network, the links may be particular objects, people or events that are common to (and hence can bridge) distinct episodic memories that include those items. But how would such memory networks be organized within the architecture of the hippocampus?

Hampson *et al.*⁶ now outline, for the first time, a functional organization of the hippocampus — a sketch of a memory network on the surface of the hippocampus. The authors monitored the activity of hippocampal neurons in rats, first when they performed a simple task, and then when they remembered the episode in a subsequent test. On each trial the animal initially pressed a 'sample' lever, presented in one of two positions in a chamber. The rat maintained this memory for several seconds, then demonstrated it by choosing the other lever when both were presented in a 'nonmatch' phase of the trial.

To make these recordings, Hampson *et al.* used a unique electrode array that allowed them to monitor many cells at known distances apart in the hippocampus (Fig. 1). In each animal the activity of some neurons was associated with the position of the lever being pressed, regardless of whether this occurred during the sample or nonmatch trial phase. Other cells fired during just one of the trial phases, independently of the lever's position. Yet other cells fired in association with the various combinations of lever position and trial phase (for example, left-match), or with several events that made up a specific type of trial (for example, right-sample then left-nonmatch). All of which means that hippocampal neuronal activity represents both the relevant aspects of space and the relevant nonspatial features of the task, consistent with the mixture of spatial and nonspatial coding observed in other situations¹¹.

By combining the data from several animals, Hampson *et al.* found a set of regular anatomical patterns (Fig. 1). Lever-position codings are segregated, such that alternating 0.6–0.8-mm cross-sectional segments of the hippocampus contain clusters of 'left' or 'right' lever-position codings. Trial-phase codings are also segregated in alternating 0.2–0.4-mm cross-sectional clusters of 'sample' and 'nonmatch' responses. The two topographies are interleaved, such that each position cluster contains clusters for both trial phases. Furthermore, in this mapping, function follows form — the clusterings of position and trial-phase specificity follow the known anatomical organization in which neurons are more closely interconnected within cross-sectional segments (called lamellae) than they are across them¹³. This correlation between anatomical and

Molecular motors

Spotting the goods trains

The organization of membranous structures in the cell, such as the Golgi body or endoplasmic reticulum, depends on motor proteins and the microtubules along which they move. Of the two major protein families involved, the dyneins move towards the 'minus' end of the microtubules, generally clustered close to the nucleus, whereas the kinesins chiefly move towards the 'plus' ends of the microtubules, distributed about the periphery of the cell.

Although organelle transport is a central function of microtubule-based motors, these motor proteins were originally identified for other reasons, such as their ability to make microtubules glide over glass coverslips. Now, however, Ron Vale and colleagues present a more direct approach (*J. Cell Biol.* 147, 493–505; 1999). By specifically assaying organelle transport they have identified two new kinesin-like motors from the slime mould



Dictyostelium discoideum.

To do this, the authors mixed *Dictyostelium* cell extracts with organelles and assembled microtubules. Then, using the video-enhanced differential interference microscopy, Vale and colleagues assessed the extracts' ability to reconstitute organelle movement, and also determined the direction of that movement. An example of this assay is shown in the picture. Video stills separated by half a second are overlaid and colour-coded to show the direction of motion (purple–cyan).

Using a number of chromatography steps, two plus-end-directed activities were purified associated with proteins of relative molecular masses

(M_r) 245,000 and 170,000. Peptide analysis allowed the identification, cloning and sequencing of the M_r 245,000 protein. It proved to be homologous to known kinesins from mice and nematode worms, and was named DdUnc104 after the worm homologue. When Vale and colleagues knocked out the *DdUnc104* gene, the mutant mould had disrupted organelle transport, confirming the *in vivo* role of DdUnc104 in this process.

The importance of this work is not so much in the identification of two previously unknown kinesins, but in the establishment of a system for dissecting the mechanics of organelle transport. Because *Dictyostelium* is so amenable to genetic manipulation, it should be possible to investigate not only the motors involved, but also the accessory proteins and receptors that allow them to identify and transport specific organelles.

Christopher Surridge

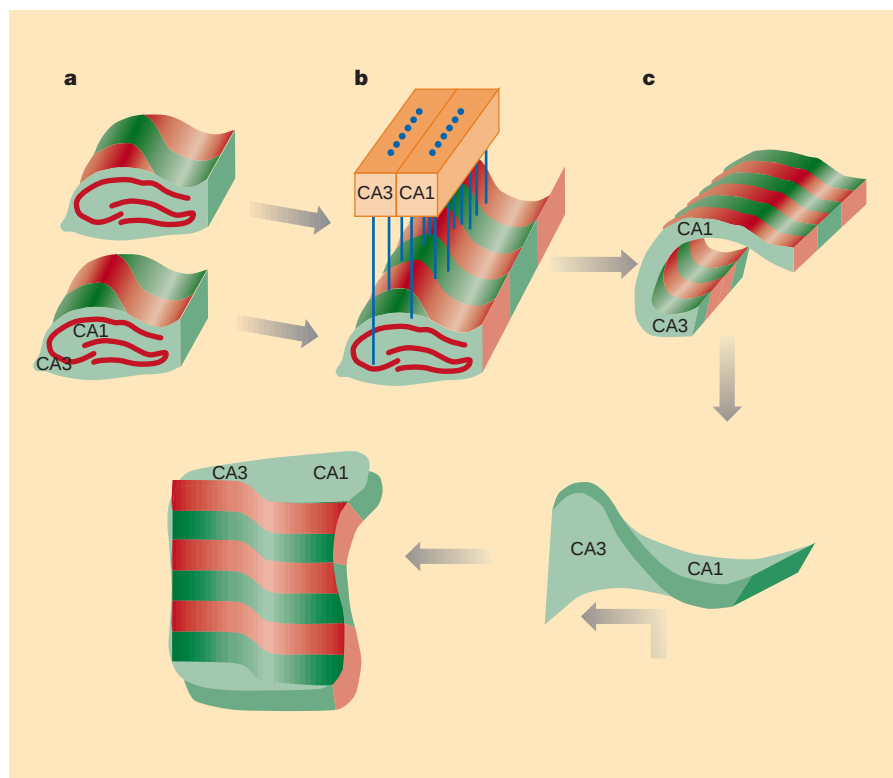


Figure 1 'Unfolding' of a functional organization in the hippocampus. **a**, Pairs of hippocampal lamellae — anatomical segments within which CA3 and CA1 are closely connected. Each pair of lamellae represents a single lever position (black or white for left or right) and a set of trial phases (sample and nonmatch). **b**, Normal appearance of the dorsal hippocampal regions CA1 and CA3, and the sites where Hampson *et al.*⁶ made recordings with the electrode array. **c**, Unfolding of the hippocampus, revealing the systematic functional architecture. (Figure provided by R. E Hampson and S. A. Deadwyler.)

functional organization resembles the common anatomical and functional 'columns' of well-connected, like-responsive cells in most areas of the cerebral cortex^{1,2}.

Two other aspects of Hampson and colleagues' observations should be emphasized. First, the representation of space in this functional organization is not a detailed Cartesian mapping, as envisaged within the dedicated spatial-mapping view of hippocampal function⁹. The authors' code involves only two positions — not a continuous set of spatial coordinates. Perhaps the hippocampus discriminates space only at the resolution required to remember where important events occurred.

Second, the full scope of declarative memory probably cannot be reduced to features of working memory such as 'position' and 'trial phase'. Yet the topographies for these features fill the hippocampus, and there are no empty slots for other features of memory. The same hippocampal cells can encode different — and seemingly unrelated — features when an animal is exposed to different situations, even in the same environment, and the scope of hippocampal coding¹² goes beyond the memory task studied by Hampson *et al.* So functional organization of the hippocampus probably contains a different set of interleaved topographies for

each of several memory networks.

Finally, the authors offer a further clue that may relate the organization of these networks to their operation. Hippocampal neu-

rons encoding combinations of the position and trial phases are found on the borders of the position and trial-phase codings. Perhaps these cells represent events unique to particular types of trial episode. The lever-position and trial-phase cells, by contrast, encode events that are common to, and might link, the representations of different episodes. A memory network based on these linked episodic codings could mediate a rat's ability to remember previous trials based on current events. Of course, this is the critical memory demand in the task. But the same kind of functional organization could mediate the linking of episodic memories — and access to them through current cues — across many domains of memory in humans as well as in animals. ■

Howard Eichenbaum is in the Laboratory of Cognitive Neurobiology, Department of Psychology, Boston University, 64 Cummington Street, Boston, Massachusetts 02215, USA.

e-mail: hbe@bu.edu

1. Mountcastle, V. B. *J. Neurophysiol.* **20**, 408–434 (1957).
2. Hubel, D. H. & Wiesel, T. N. *J. Physiol.* **160**, 106–154 (1962).
3. Gross, C. G., Roche-Miranda, D. B. & Bender, D. B. *J. Neurophysiol.* **35**, 96–111 (1972).
4. Fuster, J. M., Bauer, R. H. & Jervey, J. P. *Exp. Neurol.* **77**, 679–694 (1982).
5. Tanaka, K. *Science* **262**, 685–688 (1993).
6. Hampson, R. E., Simeral, J. D. & Deadwyler, S. A. *Nature* **402**, 610–614 (1999).
7. Cohen, N. J. & Squire, L. R. *Science* **210**, 207–210 (1980).
8. Morris, R. G. M., Garrud, P., Rawlins, J. P. & O'Keefe, J. *Nature* **297**, 681–683 (1982).
9. O'Keefe, J. A. *Exper. Neurol.* **51**, 78–109 (1976).
10. Bunsey, M. & Eichenbaum, H. *Nature* **379**, 255–257 (1996).
11. Wood, E. R., Dudchenko, P. A. & Eichenbaum, H. *Nature* **397**, 613–616 (1999).
12. Eichenbaum, H., Dudchenko, P., Wood, E., Shapiro, M. & Tanila, H. *Neuron* **23**, 1–20 (1999).
13. Andersen, P., Bliss, T. V. P. & Skrede, K. K. *Exp. Brain. Res.* **13**, 222–238 (1971).

Astrophysics

Chaotic planet formation

Renu Malhotra

Our understanding of how planets formed around the Sun is far from complete, and is increasingly challenged by new discoveries of unusual planetary-mass companions around other stars. The idea that chaotic dynamics — or extreme sensitivity to initial conditions — plays a large role in the formation, evolution and survival of planetary systems has gained popularity in recent years. This is in contrast to earlier ideas that solar systems similar to our own are an inevitable by-product of the formation of Sun-like stars.

On page 635 of this issue, Thommes *et al.*¹ propose that all of the giant planets in our Solar System formed in a narrow region of the disk-like solar nebula that surrounded the young Sun: they ended up in their present

widely spaced orbits as a result of violent and chaotic scattering. Also, on page 633, Armitage and Hansen² propose that the early formation of a Jupiter-mass planet in a nebular disk could 'trigger' the formation of other giant planets. Such protoplanets would also form in a narrow region, and would be expected to evolve chaotically into systems that might resemble some of the putative extrasolar planetary systems.

Both groups have used computer simulations to advance these ideas. In our Solar System there are two 'gas giants' (Jupiter and Saturn), having a small rocky core surrounded by a large hydrogen and helium atmosphere, and two 'ice giants' (Uranus and Neptune), which have icy mantles around their cores and only a thin atmosphere. Theorists

have struggled to explain how Uranus and Neptune formed at their present locations in the Solar System, because at such distances from the Sun the nebular disk density would have been low, and so the time required for these planets to grow is thought to be longer than the age of the Solar System.

The Thommes *et al.*¹ model postulates that the four giant planets started out as rocky cores in the Jupiter–Saturn region, and that the cores of Uranus and Neptune were violently tossed out by gravitational scattering from Jupiter and Saturn, which had already accreted large amounts of gas. In the simulations, the ejected planets remain in highly chaotic orbits for a short period of time (a few hundred thousand years), after which they settle down (by means of friction with the nebular disk) and gradually migrate out to their present, nearly circular orbits. The authors start their simulations with four cores embedded in a narrow zone of the solar nebula, and find that in 50% of their trials the giant planets end up more or less where they are now.

The model is unusual in suggesting that all four giant planets originated in the same region of the solar nebula — within a ring 5–10 astronomical units from the Sun (1 AU = Sun–Earth distance). This is substantially narrower than previous estimates of 10–20 AU for the birthplace of Uranus and Neptune³, and far inside their current orbits at 19 and 30 AU, respectively. Because the nebular disk density and orbital rotation is much higher at 5 AU than at 30 AU (ref. 4), the timescale for growing planetary cores is much reduced in this model, and so provides a solution to part of the mystery of how Uranus and Neptune formed. The solution is only partial because one still needs to explain how the Jupiter and Saturn cores accreted large amounts of gas whereas Uranus and Neptune did not.

The biggest remaining question is how realistic is the initial state. For example, it is likely that the violent character of the migration would be damped by gas-drag effects and the self-gravity of the nebular disk, both of which have been neglected in the numerical model. I find it conceivable that a process leading to the formation of a system of planetary cores embedded in a protoplanetary disk would self-regulate to remain 'marginally stable' throughout its evolution as the disk mass is gradually exhausted. This means the planets would separate relatively slowly rather than through violent dynamical instabilities as proposed by Thommes *et al.*¹. Further advances in numerical and analytical modelling of the evolution of protoplanetary disks will be needed to address this uncertainty.

Armitage and Hansen² consider the evolution of giant protoplanetary disks around other stars — disks much more massive than our solar nebula is thought to have been —

in which a giant planet forms quickly, possibly by means of a gravitational instability in the disk⁵. Using computer simulations of disk–planet interactions, the authors find that starting with a marginally stable disk, the embedded giant planet quickly clears a partial gap and also creates a spiral density pattern in the surrounding disk material. The planet accretes mass rapidly through the spiral arms, and when the planetary mass reaches four-to-five times Jupiter's mass the disk rapidly fragments around sites of orbital resonances (these are especially unstable orbits in the disk due to the planet's perturbation). Such 'triggering' of planet formation by Jupiter at its orbital resonances has been suggested previously for our Solar System⁶; the work by Armitage and Hansen represents the first demonstration of this process in a self-consistent computer simulation.

In this simulation, the disk fragments also accrete mass rapidly to become giant planets, with final masses greater than Jupiter. This process yields several giant planets in close proximity to each other, starting out in nearly circular orbits at large distances from the star (several astronomical units). The authors suggest that such a system would subsequently evolve by way of mutual gravitational scattering and eventually result in a planetary system consisting of one or more massive planets in eccentric orbits (although this latter process is not modelled in their simulation). Examples of such systems have been found in recent extrasolar planet searches^{7,8}.

As with the Thommes *et al.* model, the overriding uncertainty in these computer simulations is the plausibility of the initial state, in this case a single giant planet embedded in a massive protoplanetary disk. Are there reasonable evolutionary paths during star formation that lead to such systems? We also don't yet know whether the disk fragments produced by the simulations actually grow into planets rather than breaking up again; the simulations need to be carried out for longer (with all the necessary physics included) to address this question. Nonetheless, our understanding of planet formation is so limited that plausible new ideas are always welcome. ■

Renu Malhotra is at the Lunar and Planetary Institute, 3600 Bay Area Boulevard, Houston, Texas 77058, USA.
e-mail: renu@lpi.usra.gov

1. Thommes, E. W., Duncan, M. J. & Levison, H. F. *Nature* **402**, 635–638 (1999).
2. Armitage, P. J. & Hansen, B. M. S. *Nature* **402**, 633–635 (1999).
3. Hahn, J. & Malhotra, R. *Astron. J.* **117**, 3041–3053 (1999).
4. Lissauer, J. J. *Icarus* **69**, 249–265 (1987).
5. Boss, A. P. *Astron. J.* **503**, 923–937 (1998).
6. Patterson, C. W. *Icarus* **70**, 319–333 (1987).
7. Marcy, G. W. & Butler, R. P. *Annu. Rev. Astron. Astrophys.* **36**, 57–98 (1998).
8. <http://www.physics.sfsu.edu/~gmarcy/planetsearch/>

Daedalus

Citing to infinity

Lev Landau once proposed a logarithmic ranking for physicists. Einstein he placed in class 0.5, Dirac and Heisenberg in class 1.0, and Landau himself, modestly, in class 2.5. These days we have an objective ranking tool, the citation index. Each scientist is judged by the citations his papers gain in other scientists' papers. This criterion dominates the struggle for grants, tenure, promotion and reputation. Daedalus argues that Internet publication now permits an even greater step forward.

When all scientific papers are stored in, and accessed from, one vast computer database, it will be easy to extract all the citations attracted by a paper at any moment. Even better, the citations credited to each of the citers could also be extracted. Clearly it is more prestigious to be cited by a prolific author with many citations to his own credit, than to be cited by a citationless nobody. And these second-order citations could be judged in their turn by the quality of their third-order citers. So Daedalus is refining the business of citation scoring. His 'multi-order citation score', MOCS, will trace the citation tree of a paper to notional infinity, giving each successive order of citations a decreasing weighting.

The mathematical form of MOCS will need careful thought. It must be well-behaved, immune from strong perturbations propagated from distant anomalies in the citation tree, or instabilities from tangles of self or mutual citation. Such citations will be given, if not quite zero, at least a low weighting. The eccentric theorist who cites only himself, the author who builds endlessly on one little finding, the small, mutually back-scratching clique, cannot be allowed to gain unduly from their efforts. And logical tricks like citing a paper within itself, which might generate an infinite MOCS, will need to be identified and excluded. A multi-authored paper will divide its total MOCS among the authors, possibly in accordance with weightings declared within that paper.

Once properly optimized, MOCS will transform the struggle for scientific fame and recognition. Sadly, it will also give ammunition to those wicked sociologists who claim that science is merely a game for distributing prestige and grants among competing players.

David Jones

The further Inventions of Daedalus (Oxford University Press), 148 past Daedalus columns expanded and illustrated, is now on sale. Special *Nature* offer: m.curtis@nature.com

Coral growing on North Sea oil rigs

These installations are home to thriving colonies of an endangered cold-water coral.

This summer the coral *Lophelia pertusa* was found growing on oil platforms in the North Sea and on the Brent Spar oil-storage buoy during its decommissioning. The findings indicate that *Lophelia* has a wider distribution and a more rapid rate of growth than previously believed. The discovery also has implications for the debate over oil exploration in the Atlantic Ocean and the perceived benefits of onshore dismantling of deep-water platforms.

When Brent Spar was raised out of the sea in stages for dismantling (Fig. 1), colonies of *Lophelia* of up to 20 cm linear growth were found on the sides and bottom, which had been at depths of 60–109 m. *Lophelia* has recently been filmed at depths of 100–129 m on two platforms in the North Sea about 140 km east of the Shetland Islands. These domed colonies were up to 54 cm long, and were found on installations that have been producing oil since the late 1970s. Annual average linear growth rates of 25 mm have previously been suggested for this coral¹, but more recent estimates of 5.5–6.0 mm have been proposed². The size of the colonies on these 20-year-old platforms means that the average growth rate must have been at least 26 mm per year. If the colonies found on Brent Spar had grown at a similar rate, then *Lophelia* must have settled on the structure when it was in the Brent Field in the North Sea.

Lophelia is widespread in the Atlantic at depths of 150–1,500 m, but is particularly common on the upper continental slope at 200–600 m (refs 3,4). It has also been found at a depth of 50 m in Scandinavian fjords⁵, off the coast of Norway⁶, and on the Beryl platform⁷. Its occurrence on oil installations is the first recorded instance of live colonies of this species in the North Sea. It has been shown² that *Lophelia* occurs at temperatures of 4–8 °C. At the original Brent Spar site (61° 03' N, 01° 40' E), the water temperature at the depth where the shallowest *Lophelia* were found can exceed 10 °C (ref. 8); in the area of the two production platforms, the maximum water temperature at depths of 100–129 m varies between 7.6 and 9.7 °C.

The presence of *Lophelia* on oil-producing platforms has implications for the licensing of oil exploration. Greenpeace contends that the British Department of Trade and Industry failed to consider whether licensed oil-producing areas that may contain *Lophelia* should be designated as potential special areas of conservation and, as such, is in contravention of the



Figure 1 The top of Brent Spar being removed from the Heerema heavy-lift barge during its dismantling in Yrkefjorden, Norway.

European Commission's Species Habitat Directive (92/43/EEC).

It has been suggested that *Lophelia* is susceptible to disturbance from increased sedimentation and from the toxicological effects of drilling discharges^{9,10}. However, the apparently healthy colonies of *Lophelia* on platforms in the North Sea have been exposed to agreed quality standards of operational discharges, such as oily water, drilling muds and chemicals, and contaminants that may leak from the cuttings piles (E. Breuer, personal communication) that lie 10–15 m below some of the colonies. This indicates that it is not obviously affected by discharges from oil platforms.

Evolution

Conservation of a sex-determining gene

Vertebrates exhibit a surprising array of sex-determining mechanisms, including X- and Y-chromosome heterogametes in male mammals, Z- and W-chromosome heterogametes in female birds, and a temperature-dependent mechanism in many reptiles¹. The Y-chromosome-linked *SRY* gene initiates male development in mammals^{2,3}, but other vertebrates lack *SRY* and the genes controlling sex determination are largely unknown. Here we show that a gene implicated in human testis differentiation, *DMRT1*, has a gonad-specific and sexually dimorphic expression profile during embryogenesis in mammals, birds and a reptile (*Alligator mississippiensis*). Given the different sex-determining switches in these three groups, this gene must represent an ancient, conserved component of the vertebrate sex-determining pathway.

The human *DMRT1* gene was isolated

The occurrence of the coral raises questions about how to deal with this species, which is listed under the Convention on International Trade in Endangered Species (CITES), when platforms are decommissioned. At a meeting in Sintra in 1998 of countries belonging to the Oslo–Paris (Ospar) convention on protecting the marine environment, Ospar decision 98/3 indicated that the 'footings' of large platforms (jacket weight of more than 10,000 tonnes) might be left in place. Such an option would preserve existing colonies and might allow *Lophelia* to spread in the North Sea.

Niall Bell, Jan Smith

Cordah Environmental Management Consultants,
Kettock Lodge,
Aberdeen Science and Technology Park,
Aberdeen AB22 8GU, UK
e-mail: n.bell@cordah.co.uk

1. Mikkelsen, N., Erlenkauser, H., Killingley, J. S. & Berger, W. H. *Boreas* 5, 163–171 (1982).
2. Mortensen, P. B. & Rapp, H. T. *Sarsia* 83, 433–446 (1998).
3. Wilson, J. B. *J. Mar. Biol. Assoc. UK* 59, 149–164 (1979).
4. Zibrowius, H. *Mem. Inst. Oceanogr. Monaco* 11, 126–130 (1980).
5. Strömberg, T. *Det Kon. Nor. Vid. Skr.* 6, 1–9 (1971).
6. Hovland, M., Mortensen, P. B., Brattegard, T., Strass, P. & Rokoengen, K. *Palaios* 13, 189–200 (1998).
7. Pearce, F. *New Sci.* 164(2211), 16 (1999).
8. Turrell, W. R. & Slessor, G. *Marine Laboratory Aberdeen Report* No. 1/99 (1999).
9. Rogers, A. *New Sci.* 154(2087), 53 (1997).
10. Edwards, R. *New Sci.* 154(2085), 10 (1997).

by virtue of its homology to the *doublesex* gene from *Drosophila* and *mab-3* from *Caenorhabditis elegans*, which encode putative transcription factors with a conserved DM domain that is thought to bind DNA⁴. *DMRT1* maps to the minimal region of the small arm of human chromosome 9, which is deleted in patients with XY male-to-female sex reversal^{4–6}. A related gene, *DMRT2*, has also been mapped within this region⁷. Human genetic data indicate that *DMRT1*, alone or together with *DMRT2*, may operate in a dose-dependent fashion within the male (testis)-determining pathway. The chicken *DMRT1* homologue is located on the Z sex chromosome⁸.

We used regions mainly or totally outside the DM domain of the published human and chicken *DMRT1* complementary DNA sequences to design primers for amplification, by the polymerase chain reaction with reverse transcription (RT-PCR), of the mouse, chicken and alligator homologues. Sequence analysis confirmed the identity of the PCR products (GenBank accession numbers for mouse and alligator

DMRT1 are AF192561 and AF192560, respectively). Specific mouse and chicken PCR fragments were used as probes for whole-mount *in situ* hybridization analysis of *DMRT1* expression during urogenital development. These fragments detected single bands on Southern blots of genomic DNA, excluding the possibility of cross-hybridization with *DMRT2*. For the alligator, specific primers were used for RT-PCR analysis of whole urogenital tissues or gonads from embryos incubated at 30 °C (female-producing), 33 °C (male-producing) and 34.5 °C (female-producing).

Expression of *DMRT1* was sexually dimorphic in all three species, and was stronger in male gonads than in female ones (Fig. 1). In mouse embryos at 11.5 days post-coitum (d.p.c.), when the testis-determining *Sry* gene is activated in males and just before testicular differentiation, *DMRT1* was already being expressed in the gonads of both sexes. By 14.5 d.p.c., *DMRT1* was expressed more strongly in differentiating testes than in ovaries (Fig. 1a). Taken together with human gene results implicating *DMRT1* in XY sex reversal, these observations indicate that *DMRT1* is necessary for testis differentiation, presumably lying downstream of the master male-determinant *SRY*.

In chickens and other birds, the mechanism of sex determination is unknown. Sex may be controlled by a dominant ovary-determining gene carried on the W chromosome, or it may depend on Z-chromosome dosage (two doses for male development and one for female). *DMRT1* is Z-linked, and its expression profile implicates it in avian sex determination. As in the mouse, chicken embryos showed gonad-specific expression of *DMRT1*, with much stronger expression in developing male than in female gonads (Fig. 1b). This difference was evident before and during the time of gonadal sex differentiation (developmental stages 25–30; days 4.5–6.5 of incubation).

In birds, the Z chromosome does not seem to undergo inactivation, as the X chromosome does in mammals. Although the sexually dimorphic expression of *DMRT1* in chicken embryos may simply reflect this lack of dosage compensation, expression in the male is clearly more than twice that seen in females and suggests strong upregulation in ZZ embryos. This observation, together with the fact that expression is confined to the gonads, points to an active role for *DMRT1* in avian gonadal development. In the alligator, *DMRT1* expression was initially detectable by RT-PCR in the urogenital systems of embryos incubated at both male- and female-producing temperatures (Fig. 1c). However, gonadal expression subsequently became higher in developing male embryos than in female embryos (Fig. 1c).

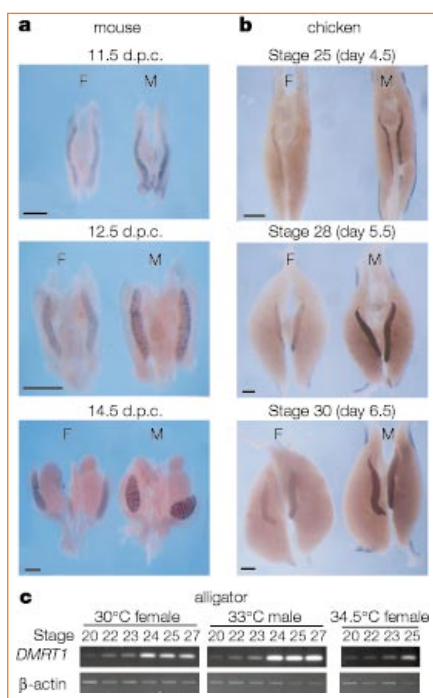


Figure 1 Sexually dimorphic expression of *DMRT1* in embryonic gonads. **a**, **b**, Whole-mount *in situ* hybridization of embryonic mouse (**a**) and chicken (**b**) urogenital systems during development, showing stronger expression (purple staining) in male than in female gonads (scale bar, 0.5 mm). **c**, RT-PCR analysis of embryonic alligator urogenital systems (developmental stages 20–23) or isolated gonads (stages 24–27) showed more upregulation of *DMRT1* expression at the male-determining temperature (33 °C) than at either female-producing temperature. β -Actin gel-loading controls are also shown.

The sexually dimorphic pattern of *DMRT1* expression in mouse, chicken and alligator embryos is consistent with a conserved role in vertebrate sex determination. We suggest that high *DMRT1* expression is necessary for testicular differentiation, whereas lower expression is compatible with ovarian differentiation. Our observations support the idea that core components of the vertebrate sex-determining pathway have been conserved during evolution. As *Drosophila* and *C. elegans* also have related genes involved in sexual development⁴, the DM-domain genes are the first to show sexually dimorphic expression across both vertebrate and invertebrate phyla.

Craig A. Smith, Peter J. McClive, Patrick S. Western, Kirsty J. Reed, Andrew H. Sinclair

Department of Paediatrics and Centre for Hormone Research, University of Melbourne, Royal Children's Hospital, Melbourne, Victoria 3052, Australia

1. Bull, J. J. *Evolution of Sex-determining Mechanisms* (Benjamin Cummings, Menlo Park, California, 1983).
2. Sinclair, A. H. et al. *Nature* **346**, 240–244 (1990).
3. Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. N. & Lovell-Badge, R. *Nature* **351**, 117–121 (1991).
4. Raymond, C. S. et al. *Nature* **391**, 691–695 (1998).
5. Fleijter, W. L. et al. *Am. J. Hum. Genet.* **63**, 794–802 (1998).
6. Guili, S. et al. *Am. J. Hum. Genet.* **63**, 905–908 (1998).
7. Raymond, C. et al. *Hum. Mol. Genet.* **8**, 989–996 (1999).
8. Nanda, I. et al. *Nature Genet.* **21**, 258–259 (1999).

Bioenergetics

Proton pumping by cytochrome *c* oxidase

Proton pumping by cytochrome *c* oxidase¹ was thought to be restricted to the oxidative part of its catalytic cycle², but this has been questioned³. New results⁴ were interpreted as an indication that two protons are pumped during the oxidative phase, and two during a subsequent reductive phase, and that this latter pumping is energetically coupled to the oxidative phase. Here I re-evaluate these results and draw an alternative conclusion.

Figure 3 of ref. 4 shows that only about 44% of the electric-field generation (caused by proton pumping, electron transfer and proton uptake) occurs during oxidation of the fully reduced enzyme. However, the claim⁴ that four protons (two during oxidation and two during re-reduction) are pumped per catalytic cycle is inconsistent with the other experiment reported at the same time (Fig. 2 of ref. 4).

There are four indications that a molar excess of oxygen over active, correctly oriented cytochrome *c* oxidase has been used (www.biophys.mpg.de/michel/public/cox-disc). For example, complete oxidation leads to pumping of only about 1.2 protons per enzyme. Re-reduction would then lead to pumping of 2.4 protons per turnover, assuming that the same number of protons were pumped during re-reduction as during oxidation. Four protons per enzyme can then be pumped, if a molar excess of oxygen is present, leading partially to a second turnover.

Assuming an optical pathway of roughly 1 cm (ref. 5) and using published spectra⁶, we calculate that the concentration of oxidized cytochrome *c* oxidase (Fig. 2a, bottom, of ref. 4) is roughly 0.70 μ M, rather than 1.25 μ M. The addition of 1.25 μ M O₂ was superstoichiometric. Another possibility is that the optical path length might have been shorter (0.56 cm), but calculations for Fig. 2 of ref. 5 indicate a longer path length (0.73 cm) for the device used.

The observation that four protons have already been pumped when only 40% of the haem groups have been re-reduced was interpreted⁴ as an indication that input of only one electron may be sufficient to elicit pumping of two protons, and was presented as support for the existence of an energy-rich state, O₂⁻. However, this result was caused by the excess of oxygen, so there is no evidence that four protons are pumped per turnover when starting from the fully reduced enzyme, or for the existence of the O₂⁻ state. The fully reduced state is not part of the natural reaction cycle of cytochrome *c* oxidase. This artificial cycle tells us little about proton pumping in the natural cycle,

especially as the fully reduced enzyme is out of electrostatic balance⁷.

The partial reversibility of the catalytic cycle by an imposed electric field led to the postulate that the transitions between the 'peroxy' (P) and 'oxoferryl' (F) intermediates, and between the F and 'oxidized' (O) intermediates, are coupled to the pumping of two protons each². O~ is now⁴ postulated to be the next intermediate after F and to have the energy to pump two protons, but to decay to the de-energized O. With the existence of O~, a reversal of O to F and then to P with equal energy inputs for pumping two protons each is impossible, as the reversal of O to F would now require the input of sufficient energy to translocate four charges. This is in contrast to the experiment represented in Fig. 3 of ref. 2, which shows that 2.2 charges are transported for the reversal from O to F.

Starting from fully reduced cytochrome *c* oxidase, a catalytic cycle in which only three protons are pumped (two during oxidation and one during re-reduction)⁷ fits the data of Fig. 3 in ref. 4 as well as a two-plus-two cycle, fits the data of Fig. 2 in ref. 4 better than a two-plus-two cycle, and avoids any need for the energy-rich state O~.

Hartmut Michel

Department of Molecular Membrane Biology,
Max-Planck-Institute for Biophysics,
Heinrich-Hoffmann-Strasse 7,
60528 Frankfurt/Main, Germany
e-mail: michel@mpibp-frankfurt.mpg.de

1. Wikström, M. *Nature* **266**, 271–273 (1977).
2. Wikström, M. *Nature* **338**, 776–778 (1989).
3. Michel, H. *Proc. Natl Acad. Sci. USA* **95**, 12819–12824 (1998).
4. Verkhovsky, M. I. et al. *Nature* **400**, 480–483 (1999).
5. Verkhovsky, M. I., Belevich, N., Morgan, J. E. & Wikström, M. *Biochim. Biophys. Acta* **1412**, 184–189 (1999).
6. Verkhovsky, M. I., Morgan, J. E. & Wikström, M. *Biochemistry* **34**, 7483–7491 (1995).
7. Michel, H. *Biochemistry* (in the press).

Verkhovsky et al. reply — Ten years ago, one of us showed that all proton translocation by cytochrome *c* oxidase seems to be coupled to the enzyme's oxidative catalytic half-cycle¹. This led to the assumption that all proton translocation takes place during this half of the cycle. Our more recent data² contradict that idea, as well as Michel's hypothesis³, but our original findings¹ still stand.

The main reason for our proposal that proton translocation during the reductive half of the cycle may still be energetically coupled to the preceding oxidative phase was our observation that, after the energy-rich O~ state had relaxed to O, there was no proton translocation associated with reduction². Michel does not mention this, although it contradicts his model³.

Michel is right that our² Fig. 2a shows the release of about 1.2 H⁺, but this was in a single experiment. In the text² we reported the release of 1–2 H⁺. This variation is due to the difficulty of having

all the enzyme molecules reduced by exactly four electrons. When more than four electrons are available (Fig. 2b of ref. 2), there is no such difficulty and the enzyme is always fully reduced. The result in Fig. 2b therefore shows much less variation, and consistently yielded four released H⁺. In the single trace of Fig. 2a, all enzyme molecules had not been fully reduced by four electrons. This can be seen from the small, slow tail in the corresponding absorbance trace, resulting from the slow decay of oxygen intermediates that arise when the fraction of enzyme with less than four electrons reacts with oxygen.

All the enzyme relevant to these experiments is correctly orientated relative to the membrane. We ascertained that incorrectly orientated enzyme is not reduced under our experimental conditions by the charged reductants ferrocyanide *c* and hexammine ruthenium [II].

It has been shown⁴ that, in conditions similar to those used for our Fig. 2b (ref. 2), any additional turnovers resulting from an excess of oxygen over the enzyme would be very slow (roughly 15 seconds), as such turnovers depend on slow intermolecular reactions. Because the reaction in Fig. 2b is much faster, Michel's claim that the result is due to multiple turnovers cannot be correct. Michel's conclusion that the O₂ concentration was superstoichiometric in this experiment is also flawed, because we ascertained the stoichiometry of O₂ versus enzyme by titrating the reduced enzyme spectroscopically with known amounts of O₂. Thus, both the enzyme concentration and the number of protons released were determined on the basis of the concentration of O₂ added.

Michel's comments about extinction coefficients are therefore not relevant, and they are also incorrect. He concludes that the path length was ~1 cm in our experiments², but we stated⁵ that the path length varies from experiment to experiment depending on the optical geometry, and that it must be determined separately each time. A calculation based on Fig. 2 of ref. 5 would have shown this.

Finally, assuming there would have been both an excess of O₂ and an excess of reductant (cytochrome *c*) in the experiment² shown in our Fig. 2b, and, as claimed by Michel, the first turnover would pump three protons and the subsequent turnovers four protons each, we can calculate that nine turnovers would be needed to yield 3.9 H⁺ per O₂ (note that H⁺ per O₂ is the primary quantity measured). Michel calculates 1.8 turnovers, which would give only 3.4 protons, far below the number observed. Obtaining these extra turnovers would require not only an excess of O₂, but also an excess of ferrocyanide *c*. Our experiment had neither of these but, even if

it had both, our result could still not be explained by multiple turnovers.

**M. I. Verkhovsky, A. Jasaitis,
M. L. Verkhovskaya, J. E. Morgan,
M. Wikström**

Department of Medical Chemistry,
Institute of Biomedicine, University of Helsinki,
00014 Helsinki, Finland

1. Wikström, M. *Nature* **338**, 776–778 (1989).
2. Verkhovsky, M. I. et al. *Nature* **400**, 480–483 (1999).
3. Michel, H. *Proc. Natl Acad. Sci. USA* **95**, 12819–12824 (1998).
4. Antonini, E., Brunori, M., Greenwood, C. & Malmström, B. G. *Nature* **228**, 936–937 (1970).
5. Verkhovsky, M. I., Belevich, N., Morgan, J. E. & Wikström, M. *Biochim. Biophys. Acta* **1412**, 184–189 (1999).

Microporous materials

Electrochemically grown photonic crystals

There is a particular group of microporous materials (dielectric structures that exhibit periodicity in three dimensions, called photonic crystals) that are of special interest because of their unique optical properties¹: they can manipulate light in much the same way that a superconductor manipulates electrons. Here we use a technique involving electrochemical deposition within a colloidal template to create porous materials that are stronger and have a higher refractive index than has previously been possible.

In the quest to increase refractive-index contrast in microporous materials, attempts have been made to use colloidal assemblies as templates in a replication process^{2–5}. The feasibility of this approach is supported by calculations predicting that a matrix having a high index of refraction and containing a face-centred cubic lattice of spheres of low refractive index, notably air, can give rise to a complete gap in the photonic band structure^{6,7}. Although such structures can be built up layer by layer through conventional lithography, processing difficulties limit the formation of more than a few layers^{8,9}.

Structures that have substantial three-dimensional periodicity are required for many photonic applications, but so far it has proved difficult to fully mineralize the interstitial space of a colloidal assembly and generate such a three-dimensional structure. In principle, this involves only two steps: the mineralization of the interstitial space of a colloidal assembly, and the subsequent removal of the initial colloid template (and possibly some by-products of chemical reactions). Previous promising attempts have produced replicas with only moderate density and thus an effective index of refraction that is smaller than that of the bulk material. They have also been mechanically fragile^{2–5,10}.

Colloidal assemblies are attractive because of their size tuneability and three-

dimensional structure. A critical issue in creating photonic materials is how to fill the interstitial space completely with a component that has a high refractive index. We therefore electrodeposited group II–VI semiconductors into colloidal sediments, filling the interstitial space with materials of high refractive index (Fig. 1). The electrodeposited materials still allow the colloidal template to be removed, which is necessary to maximize the refractive-index contrast.

Electrodeposition should be an ideal way to fill topologically complex structures because it starts from deep within the structure and then grows out towards the exposed surfaces. We obtained microporous cadmium–selenium (CdSe) structures by potentiostatic¹¹ deposition in the interstitial space of a polystyrene colloidal template (Fig. 1a,b). Cadmium sulphide (CdS) was also grown by galvanostatic¹² (Fig. 1c) and potentiostatic (results not shown) deposition in colloidal assemblies, again yielding three-dimensionally periodic structures after the template is removed.

Because both CdS and CdSe have high refractive indices (2.5 at 600 nm and 2.75 at 750 nm, respectively), these structures can have deep gaps in their photonic band structure after the template is removed. CdS and CdSe are promising materials for photonic applications because, unlike most other high-refractive-index materials, they are optically transparent in the visible and near-infrared region of the spectrum.

One of the critical requirements for the existence of a true omnidirectional photonic band gap is three-dimensional periodicity and uniformity. Our template-directed electrodeposition of semiconductors results in three-dimensional structures (Fig. 1b,c), as indicated by the clearly visible holes into the layer below in all the systems. These holes form at contact points between the spheres of the template, and their arrangement on a triangular lattice is evidence that the hollow voids are on a hexagonal close-packed array.

As well as determining the three-dimensional structure, the holes between the voids allow for the complete removal of the template by solvent or acid, or through burnout. Our electrodeposited materials appear to be dense, as observed in high-magnification scanning electron micrographs of other template-directed electrodepositions¹¹. As a result, the microporous structures shrink by only a tiny amount (<2%) when the template is removed.

Template-directed electrodeposition is well suited to creating three-dimensional microstructures: many materials of high refractive index can be electrodeposited, including semiconductors of groups II–VI, III–V and IV, all of which are very reluctant to form three-dimensional microstructures by traditional techniques. These materials, which have many desirable optical and

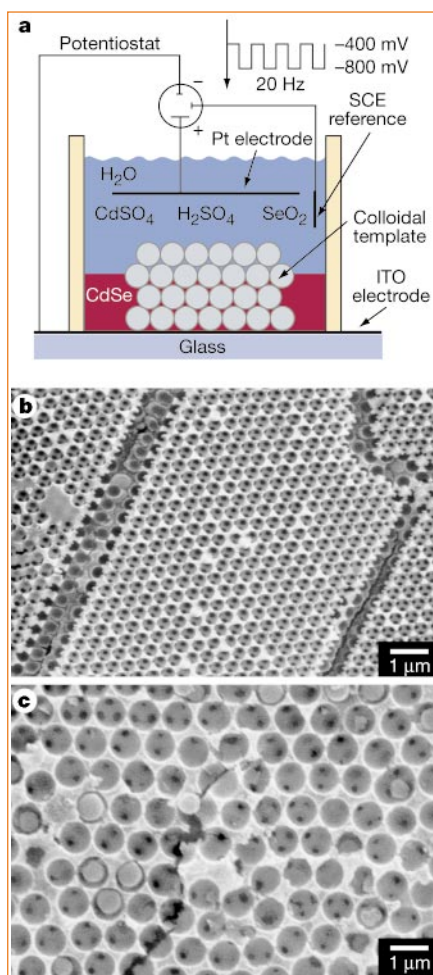


Figure 1 Electrodeposition in colloidal assemblies. **a**, Schematic representation of the experimental set-up. **b**, **c**, Scanning electron micrographs of the product. **b**, CdSe potentiostatically deposited in the interstitial space of a polystyrene colloidal assembly formed from polystyrene spheres 0.466 μm in diameter. After the electrodeposition, the polystyrene template was fully dissolved with toluene. SCE, standard calomel electrode; ITO, indium tin oxide. **c**, CdS galvanostatically grown around a silica colloidal template formed from silica spheres 1 μm in diameter. The template was partly dissolved after electrodeposition by dipping the sample into an aqueous 4.8% hydrogen fluoride solution for 10 min. This did not completely dissolve the template, allowing the relation between the colloidal template and the resulting three-dimensional structure to be seen directly.

electronic properties, are commonly used in the semiconductor industry. The potential offered by integrating photonic band-gap structures with semiconductor devices could revolutionize optoelectronics and optical computing. It will be straightforward to extend our methods to metallic materials, which can readily be electrodeposited, and these three-dimensionally microperiodic metallic structures could well have unusual mechanical or thermal properties.

Paul V. Braun*, **Pierre Wiltzius**

Bell Laboratories, Lucent Technologies,
700 Mountain Avenue, Murray Hill,
New Jersey 07974, USA

*Present address: Department of Materials Science

and Engineering, 1304 West Green Street,
Urbana, Illinois 61801, USA
e-mail: p-braun@uiuc.edu

1. Joannopoulos, J. D., Meade, R. D. & Winn, J. N. *Photonic Crystals: Molding the Flow of Light* (Princeton Univ. Press, New Jersey, 1995).
2. Holland, B. T., Blanford, C. F. & Stein, A. *Science* **281**, 538–540 (1998).
3. Wijnhoven, J. E. G. J. & Vos, W. L. *Science* **281**, 802–804 (1998).
4. Zakhidov, A. A. *et al. Science* **282**, 897–901 (1998).
5. Vlasov, Y. A., Yao, N. & Norris, D. J. *Adv. Mater.* **11**, 165–169 (1999).
6. Biswas, R., Sigalas, M. M., Subramania, G. & Ho, K.-M. *Phys. Rev. B* **57**, 3701–3705 (1998).
7. Busch, K. & John, S. *Phys. Rev. E* **58**, 3896–3908 (1998).
8. Lin, S. Y. *et al. Nature* **394**, 251–253 (1998).
9. Fleming, J. G. & Lin, S. Y. *Opt. Lett.* **24**, 49–51 (1999).
10. Imhof, A. & Pine, D. J. *Nature* **389**, 948–951 (1997).
11. Klein, J. D. *et al. Chem. Mater.* **5**, 902–904 (1993).
12. Edamura, T. & Muto, J. *Thin Solid Films* **235**, 198–201 (1993).

addendum

Colour categories in a stone-age tribe

J. Davidoff, I. Davies, D. Roberson

Nature **398**, 203–204 (1999)

The colour-naming pattern of Berinmo speakers may appear consistent with that obtaining in tritanopia¹, a colour-vision disorder that has an increased frequency in other tropical areas² and can arise from chronic exposure to short-wavelength light³. Furthermore, it could be argued that an isolated tribe may have become tritanopic through a shared genetic defect. However, our Berinmo speakers were not tritanopic. They were tested with the City University Colour Vision Test⁴ that specifically assesses tritanopia. The test consists of ten plates, and all speakers who failed any of the plates were eliminated from our study. The failure rate was 7 out of 83 speakers tested during the course of the study.

1. Pokorny, J., Smith, V. C., Verriest, G. & Pinckers, A. J. L. G. *Congenital and Acquired Color Vision Defects* (Grune & Stratton, New York, 1979).
2. Davies, I. R. L., Laws, G., Corbett, G. G. & Jerrett, D. J. *Pers. Ind. Diff.* **25**, 1153–1162 (1998).
3. Werner, J. S., Peterzell, D. H. & Scheetz, A. J. *Optom. Vis. Sci.* **67**, 214–229 (1990).
4. Fletcher, R. *The City University Colour Vision Test* 2nd edn (Keeler, London, 1980).

corrections

Magnet levitation at your fingertips

A. K. Geim, M. D. Simon

Nature **400**, 323 (1999)

We inadvertently omitted a second reference to the work of I. A. H. Boerdijk (*Phillips Res. Rep.* **11**, 45–56; 1956), where the first demonstration, to our knowledge, of levitation of a magnet using non-superconducting diamagnetic materials as stabilizers was reported. The more stable two-sided configuration was first described by Ponisovskii (*Prib. Tekh. Eksp.* **24**, 7–14; 1981). We were unaware of the earlier work until our experiments were completed.

Tropical tree gene flow and seed dispersal

M. B. Hamilton

Nature **401**, 129–130 (1999)

Two numbers were transposed in the last column of Table 1: for site 7, the values for populations 41–2 and 3209 should have been 1.0 and 0.0, respectively.

The role of stress transfer in earthquake occurrence

Ross S. Stein

US Geological Survey, MS 977, Menlo Park, California 94025, USA

An earthquake alters the shear and normal stress on surrounding faults. New evidence strengthens the hypothesis that such small, sudden stress changes cause large changes in seismicity rate. Rates climb where the stress increases (aftershocks) and fall where the stress drops. Both increases and decreases in seismicity rate are followed by a time-dependent recovery. When stress change is translated into probability change, seismic hazard is seen to be strongly influenced by earthquake interaction.

During the 75 years before the great 1906 earthquake on the San Andreas fault, the San Francisco Bay area suffered at least 14 shocks of moment magnitude (M_w) equal to or exceeding 6; these occurred on all major faults, and included two events of $M_w \geq 6.8$. In the succeeding 75 years, there was but one $M_w \geq 6$ shock¹ (Fig. 1). Elsewhere, $M_w \geq 6$ earthquakes in the extensional regime seaward of subduction zones occur, with few exceptions, only in the years following great subduction events². Evidently, the rate of seismicity is therefore not constant, and the rate—or probability—of earthquakes on one fault is not independent of the rate on another. Yet there is nothing in probabilistic seismic hazard assessment (the principal tool of the engineering, insurance, financial, and emergency-response communities) that reflects or can reproduce such observations. Earthquake interaction is a fundamental feature of seismicity, leading to earthquake sequences, clustering, and aftershocks. One interaction criterion that promises a deeper understanding of earthquake occurrence, and a better description of probabilistic hazard, is Coulomb stress transfer.

Coulomb failure stress

An earthquake reduces the average value of the shear stress on the fault that slipped, but as Chinnery first showed in 1963, shear stress

risks in more areas than just the fault tips³. The importance of this discovery was realized about 20 years later, when lobes of off-fault aftershocks were seen to correspond to small calculated increases in shear⁴ or Coulomb stress^{5,6}. In its simplest form, the Coulomb failure stress change, $\Delta\sigma_f$ (also written ΔCFS or ΔCFF) is

$$\Delta\sigma_f = \Delta\tau + \mu(\Delta\sigma_n + \Delta P) \quad (1)$$

where $\Delta\tau$ is the shear stress change on a fault (reckoned positive in the direction of fault slip) and $\Delta\sigma_n$ is the normal stress change (positive if the fault is unclamped). ΔP is the pore pressure change in the fault zone (positive in compression), and μ is the friction coefficient (with range 0–1). Failure is encouraged if $\Delta\sigma_f$ is positive and discouraged if negative; both increased shear and unclamping of faults promote failure. The tendency of ΔP to counteract $\Delta\sigma_n$ is often incorporated into equation (1) by a reduced 'effective' friction coefficient, μ' (ref. 7).

The calculated off-fault stress increases are rarely more than a few bars (1 bar = 0.1 MPa, which is approximately atmospheric pressure at sea level), or just a few per cent of the mean earthquake stress drop. In addition, the proximity to failure at any site is presumably variable but in any event unknown. So why would aftershocks concentrate at the site of such small stress increases? New studies



Figure 1 Comparison of earthquakes before and after the $M_w = 7.8$ San Francisco earthquake on the San Andreas fault. Solid red lines, interpreted rupture positions⁴²; dashed red lines, the 1906 earthquake. Urban areas are shown grey. S.F. Bay, San Francisco Bay. Although this is the longest historical earthquake record in the western

United States, it is probably complete for $M_w \geq 6$ only since the 'gold rush' of 1849, and so underestimates the rate of shocks during the pre-1906 period. The southern end of the 1906 rupture lies near the bottom of the image; the northern end lies 200 km northwest of the image. Processing by R. E. Crippen.

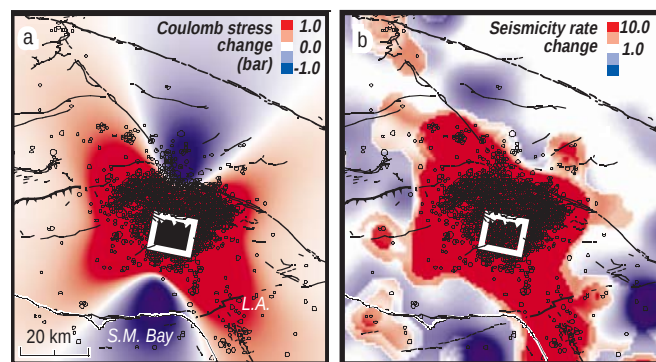


Figure 2 Correlation between calculated Coulomb stress change and seismicity rate change for the 1994 $M_w = 6.7$ Northridge earthquake. **a**, The largest Coulomb stress change on optimally oriented thrust or strike-slip faults at depths of 3–10 km; the compressive axis of the regional stress is oriented $N4^\circ E$, and $\mu = 0.4$ (ref. 43). Locations of active surface faults, and $M_w \geq 1.5$ shocks during 3–6 months after the mainshock, are superimposed in black. **b**, The seismicity rate change (new/old), comparing the rate during the 78 months before the Northridge earthquake to that 3–6 months afterwards⁴⁴ (the first 3 months after Northridge have a poorer level of completeness and are excluded). The rate is calculated in 10-km cells on a grid with 1-km spacing and then smoothed with a gaussian filter. The rate change in the white areas is unresolved. About 65% of the resolved area is positively correlated. Such correlations extend 40 km (or 4 fault lengths) southeast of the mainshock to Los Angeles (L.A.). Observed seismicity rate decreases in the Santa Monica Bay (S.M. Bay) and along parts of the San Andreas fault are correlated with the calculated stress decrease.

find a surprisingly strong influence of stress change on seismicity, and seek to explain it in terms of rupture nucleation phenomena observed in the laboratory.

Stress change and seismicity rate change

More than any other earthquake, the 1992 $M_w = 7.3$ shock in Landers, California, changed the landscape of stress-triggering investigations. Rich in aftershocks, this well-recorded event enabled detailed estimates to be made of the distribution of the fault slip, needed to calculate the stress changes. Because in map view the strike-slip rupture is concave to the west, the earthquake was calculated to produce a 2-bar lobe of Coulomb stress increase 40 km west of the mainshock, where the $M_w = 6.5$ Big Bear shock struck 2.5 hours after Landers⁸. While this association alone could result from chance, 67% of the 10,000 $M > 1$ Landers–Big Bear aftershocks also occurred in regions calculated to have been brought >0.1 bar closer to failure (termed the stress-triggering zones), and few off-fault aftershocks occurred in regions inhibited by >0.1 bar from failure (the stress shadows)^{7,9–11}. Most of these comparisons of stress change to aftershocks rely on the assumption that small shocks occur on planes optimally oriented for failure as a result of the regional stress and the earthquake stress change⁷. The association of calculated Coulomb stress increases with aftershocks is now widely reported (see ref. 12 and references therein).

Tantalizing as it may be, finding aftershocks in the stress trigger zones does not demonstrate that the stress imparted by the mainshock had any effect on off-fault seismicity, as seismicity may have been as abundant in those zones before the mainshock. A stronger test of the Coulomb hypothesis is to compare the calculated stress change to the observed seismicity rate change¹³. After a mainshock, sites of both increased and decreased seismicity rate are seen; for the $M_w = 6.7$ 1994 Northridge, California, shock, 65% of the observed seismicity rate changes are correlated with the calculated Coulomb stress change (Fig. 2).

An earthquake can thus enhance or suppress subsequent events, depending on their location and orientation. Viewed in this light, aftershocks are simply sites of seismicity rate increase, occurring

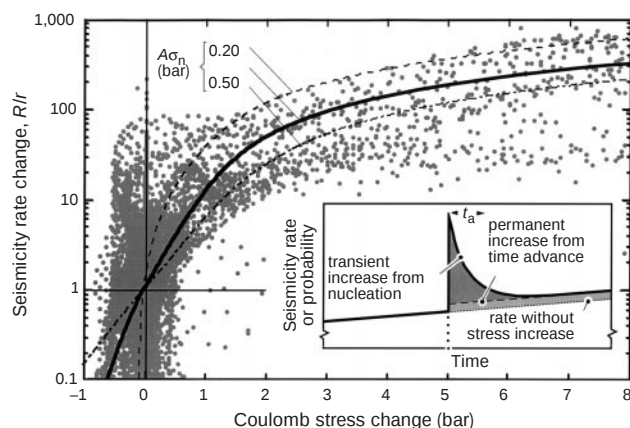


Figure 3 Observed seismicity rate change as a function of calculated Coulomb stress change for the 1995 $M_w = 6.9$ Kobe earthquake³⁴. Main figure, seismicity rate change for $M_w \geq 2.6$ earthquakes (the minimum magnitude for which the catalogue is complete) during 8 years before the Kobe shock compared to the following 1.5 years; the seismicity rate increased after the Kobe earthquake where the rate change $R/r > 1$ and fell where $R/r < 1$. The maximum Coulomb stress change on optimally oriented strike-slip and thrust faults is calculated at depths of 0–20 km, for $\mu = 0.4$. Inset, schematic illustration of the time-dependence of seismicity rate (graphically similar to conditional earthquake probability), following a sudden stress increase; the transient effect of the stress increase decays to the permanent rate change over the aftershock duration, t_b (refs 33, 34). $A\sigma_n$ is the product of the state/rate constitutive parameter A (ref. 30) and the total normal stress acting on the fault, σ_n .

where the stress has increased—whether on the fault rupture or off. Sites of seismicity rate decrease, or where the rate was higher before the earthquake than after, might logically be called ‘antishocks’ (in the sense of ‘antipasto’—they precede rather than follow the main course). A spatial regression of stress change on seismicity rate change for the 1995 $M_w = 6.9$ Kobe, Japan, earthquake (Fig. 3, main panel) reveals just how strong this effect is: a 1-bar stress increase corresponds to a 10-fold increase in rate of shocks with local magnitude $M_L \geq 2.6$; a 5-bar stress change is associated with a 100-fold rate increase. But are such findings valid if small aftershocks, whose nodal planes are unknown, do not occur on faults optimally oriented for failure? One alternative is to consider only seismicity on (say, within 1 km of) major active faults and assume that these shocks have the same strike, dip and rake as the fault on which they occur. Another approach is to calculate the Coulomb stress change on the nodal planes of the subset of shocks with known focal mechanisms. Both approaches yield new insights, as outlined below.

Several researchers^{13–15} have examined stress changes and seismicity on major faults within 100 km of the 1989 $M_w = 6.9$ Loma Prieta, California, shock. Parsons *et al.*¹⁵ found that the seismicity rate change is associated with the calculated shear stress change for major faults (slip rates more than ~ 7 mm yr^{-1} and cumulative slip of more than ~ 50 km; Fig. 4a). For minor faults (with negligible cumulative slip and lower slip rates—typically thrust or oblique faults), seismicity is concentrated where the faults were unclamped (Fig. 4b); both correlations are statistically significant. Restated in terms of equation (1), the effective friction coefficient on the major faults is low ($\mu' \leq 0.2$), while for minor faults it is high ($\mu' \geq 0.8$). This inference accords with independent arguments that major faults such as the San Andreas develop thick, impermeable gouge zones that reduce sliding friction or trap pore fluids^{16,17}, both of which lower μ' . The 1997 sequence of eight shocks, all of $M_w = 5–6$, that occurred in Umbria-Marche, Italy, on normal faults with low slip rates and modest net slip, also appears to have been promoted by unclamping¹⁸. Seeber and Armbruster¹⁹ found that the ratio of encouraged to discouraged aftershocks of the Landers earthquake is

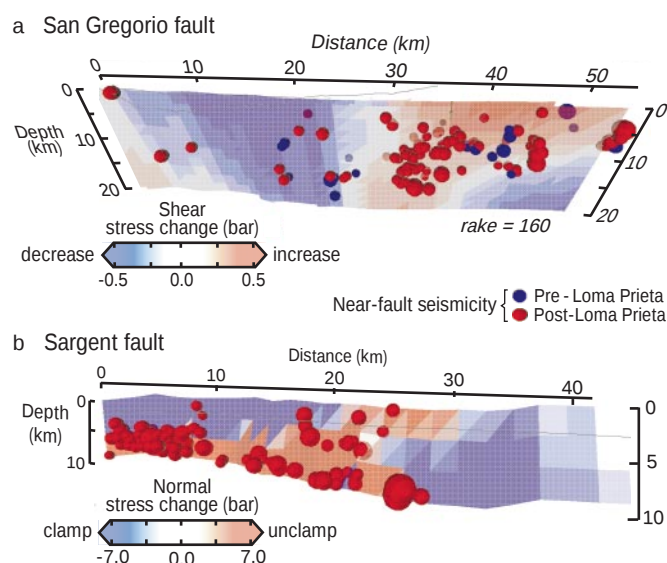


Figure 4 Seismicity and stress changes associated with the 1989 $M_w = 6.9$ Loma Prieta earthquake resolved on nearby faults¹⁵. Fault locations are shown in Fig. 6a. Seismicity with $M_L \geq 1.5$ during 7 years before and after Loma Prieta is plotted, with size proportional to magnitude. Distance increases towards the southeast. **a**, The fault is assumed to dip 70° northeast with a 160° rake, consistent with focal mechanisms and marine terrace deformation; the thin line is the coast. Earthquakes within 2 km of the fault are shown. The change in the seismicity distribution is associated with the calculated shear stress change, suggesting that μ is low (no correlation is seen for normal stress change). **b**, The fault is assumed to dip $55 \pm 5^\circ$ southwest and has a 135° rake; the thin line is the surface projection of the San Andreas fault. Earthquakes within 1 km of the fault are shown. The post-1989 seismicity is concentrated where the fault was unclamped, suggesting a high μ (no correlation is seen for shear stress change).

greatest if $\mu' = 0.85$; minor faults surround Landers. Thus the high friction inferred for minor faults seems to be borne out by several Coulomb studies.

Using focal mechanisms to evaluate the Coulomb hypothesis affords additional insights but presents new problems, because the Coulomb stress change is different on the two nodal planes of the focal mechanism, unless $\mu = 0$. Hardebeck *et al.*²⁰ calculated the Coulomb stress change on both planes, and examined whether the percentage of encouraged planes increases after the Landers and the Northridge earthquakes (Fig. 5a); the period before the earthquakes served as a control group. The 20–25% increase in the percentage of encouraged shocks (Fig. 5c and d) is statistically significant for stress increases ≥ 0.1 bar, and persists for the 5 years examined after both mainshocks. Seeber and Armbruster¹⁹ interpreted the fault plane for each of the 1900 aftershock focal mechanisms of Landers, and reached similar conclusions, although they found a larger initial increase in encouraged aftershocks that decays with time (Fig. 5b). A test for aftershocks of the 1987 $M_w = 6.6$ sequence at Superstition Hills, California, also shows significant correlations for stress increases ≥ 0.1 bar during 1.4–2.8 yr after the mainshock²¹.

Dynamic and tidal Coulomb stress change

The seismic waves excited by earthquakes produce dynamic Coulomb stress changes that, at distances more than about one source dimension from the fault, can be an order of magnitude larger than the static stress changes. How can one distinguish whether the static or dynamic stresses control seismicity? In other words, is it strong shaking or the weak permanent stress changes that promote seismicity? Because the dynamic stresses oscillate, they are everywhere positive at some point in time. All sites are shaken, and thus the dynamic stresses cast no stress shadows and should produce no 'antishocks' (seismicity rate decreases), at odds with observations.

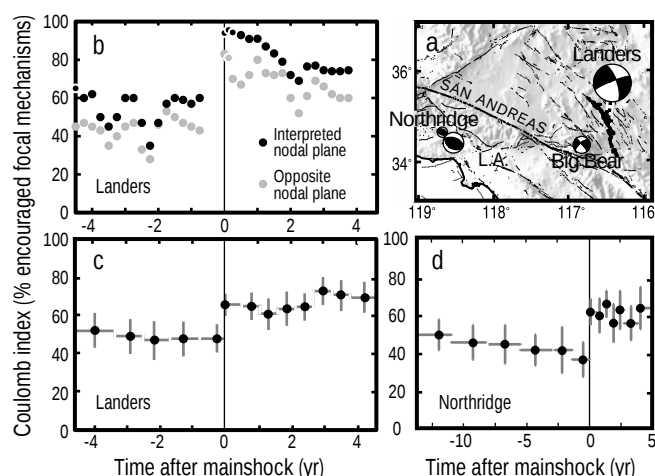


Figure 5 Stress changes associated with the 1992 $M_w = 7.3$ Landers and the 1994 $M_w = 6.7$ Northridge ruptures resolved on nodal planes of earthquakes with focal mechanisms. The percentage of planes brought closer to failure ('encouraged') is $\sim 50\%$ by chance before the main ruptures, but increases significantly afterwards, indicating that aftershocks tend to occur where the main rupture has promoted Coulomb failure. **a**, Focal mechanisms of the mainshock and largest aftershock in both sequences; LA, Los Angeles. **b**, $M_L \geq 1.5$ shocks ≥ 27 km from the Landers rupture are used, at sites where the calculated stress change is ≥ 0.2 bar, assuming that $\mu = 0.85$ (ref. 19). **c**, $M_L \geq 2.5$ shocks within a box defined by $32.5\text{--}36.5^\circ \text{N}/115.5\text{--}119.5^\circ \text{W}$ are included, assuming $\mu = 0.4$ (ref. 20); 95% confidence bounds are shown. Events with both planes encouraged receive more weight than those with one encouraged plane. **d**, $M_L \geq 2.5$ shocks within a box defined by $33.8\text{--}34.9^\circ \text{N}/117.8\text{--}119.5^\circ \text{W}$, assuming $\mu = 0.4$ (ref. 45); 95% confidence bounds shown.

Belardinelli *et al.*²² calculated the dynamic stress evolution in the 1980 $M_w = 6.9$ Irpinia, Italy, sequence in which nearby faults ruptured 20 s apart. The second event was not triggered at the time of the dynamic peak. Rather, a delayed triggering mechanism must be involved irrespective of whether static or dynamic stresses are responsible, because the second rupture nucleated 12 s after the dynamic peak and 6 s after the static value had been reached.

Other evidence, however, suggests that at larger distances from the rupture, dynamic stresses may explain the distribution of seismicity rate changes better than the static stresses. Kilb *et al.*²³ found that the pattern of dynamic Coulomb stress changes bears similarities to that for static stress changes. Although the peak dynamic stress field lacks shadows, it does exhibit lobes with small stress change in roughly the same positions as the static stress shadows. The vital difference is that the dynamic stress increases are an order of magnitude larger in the direction of rupture propagation. The Landers rupture propagated unilaterally to the northwest, and produced more aftershocks in this direction²⁴. The observed seismicity rate may thus be influenced by both static and dynamic effects.

If static stress changes influence earthquake occurrence, then seismicity rates might be modulated by the solid Earth tides, the distortion of the Earth caused by the pull of the Sun and Moon. Unlike earthquakes, the tides produce no strong motion (shaking), but they do alter the stress on faults. The tidal Coulomb stress range, dominated by the normal-stress component, is only about ± 0.01 bar, or one-tenth of the threshold of detection in the most sensitive aftershock studies. Vidale *et al.*²⁵ calculated the tidal stresses on the fault planes of 13,000 earthquakes along the creeping portions of the San Andreas and Calaveras faults, and found that the seismicity rate is higher at times when the tidal stresses unclamped the fault, but not significantly so. Lockner and Beeler²⁶ cycled stress

in a laboratory sample to simulate the tides, and found that stress changes ≥ 0.1 bar caused strong correlations in the timing of stick-slip events, in accord with aftershock studies. They estimated that if detection increased with the square root of the sample size, more than 20,000 earthquakes would be needed to find a statistically significant association with tidal stresses, in which case $\sim 1.5\%$ of the seismicity would be correlated. Vidale *et al.*²⁷ repeated their experiment with 27,500 quakes, and found that the rate of seismicity during the peak tidal unclamping is 1.0% higher than average, a difference significant at the 95% confidence level. Thus the tides perceptibly alter the rate of seismicity, suggesting that the much larger off-fault stress changes associated with earthquakes are indeed one cause of seismicity rate changes.

Incorporating stress transfer into probabilistic hazard

The simplest way to incorporate stress transfer into probability models is to assume that a sudden stress change will alter the time until the next large earthquake by the ratio of the stress change on the fault to its long-term stressing rate. This is the 'time advance or delay' used in some consensus probability forecasts^{28,29}. Because stress changes on nearby faults are typically of the order 1 bar, and stressing rates are of the order 0.1 bar yr^{-1} , inter-event times are only changed by decades. Such a time change is inevitably much smaller than the uncertainty or variability of the earthquake inter-event time (typically assigned to be $\pm 50\%$), and thus has little effect on the probability. But why, then, would earthquake stress changes exert such a strong influence on seismicity rates? The 1906 stress decrease on the faults in the San Francisco Bay area, for example, is a few bars, but the rate of $M_w \geq 6$ shocks dropped by at least an order of magnitude during the ensuing 75 years (Fig. 1), in a manner consistent with the Kobe results (Fig. 3a).

A way out of this problem is suggested by the concept of state and rate friction. The observed dependence of seismicity rate on Coulomb stress change (Fig. 3, main panel) is well described by Dieterich's earthquake-rate relation^{30,31}. In state and rate friction, seismicity is viewed as a sequence of independent nucleation events in which the 'state' depends on the fault slip, slip rate, and elapsed time since the last event. In the absence of a stress perturbation, the seismicity rate is constant. But under the assumption that there is a large number of earthquake nucleation sites on the fault, there is a nonlinear dependence of the time to instability on stress change. The 'time advance' causes a modest but permanent increase in earthquake rate and probability. The transient effect of the stress change strongly amplifies the permanent change, because the fault slips at a higher rate, causing a higher rate of earthquake nucleation (Fig. 3, inset). The transient effect decays as the supply of nucleation sites is consumed; the duration of the transient is inversely proportional to the fault stressing rate.

The seismicity rate equation³⁰ in simplest form is

$$R(t) = \frac{r}{\left[\exp\left(\frac{-\Delta\sigma_f}{A\sigma_n}\right) - 1 \right] \exp\left(\frac{-t}{t_a}\right) + 1} \quad (2)$$

in which R is the seismicity rate as a function of time, t , following a Coulomb stress change, $\Delta\sigma_f$. A is a constitutive parameter, σ_n is the total normal stress, t_a is the aftershock duration (equal to $A\sigma_n/\dot{\tau}$, where $\dot{\tau}$ is the stressing rate on the fault), and r is the seismicity rate before the stress perturbation. To evaluate equation (2), the Coulomb stress change is calculated and r , t_a and $\dot{\tau}$ are estimated from observations, permitting $A\sigma_n$ to be inferred (Fig. 3, main panel).

The rate equation (equation (2)) has the form of Omori's law, which describes the observed temporal decay of aftershocks on the mainshock rupture surface. Thus the decay of seismicity following a stress change may instead be a general property of earthquakes, restricted neither to aftershocks nor to the rupture surface. Such an interpretation is borne out by observations: even at distances more

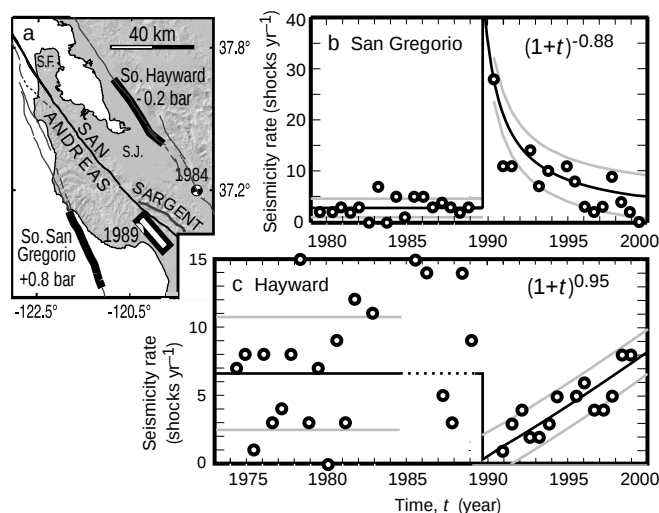


Figure 6 The influence of stress changes associated with the Loma Prieta earthquake on $M_L \geq 1.5$ seismicity rates³². Such influence is found well outside what is traditionally regarded as the aftershock zone. **a**, Map of Loma Prieta source (rectangle) and Southern Hayward and southern San Gregorio fault sections (bold) on which stress change is calculated and seismicity rates are measured: S.F. is San Francisco; S.J. is San Jose. **b**, Seismicity rate jumps on the southern San Gregorio fault, 40 km west of Loma Prieta. The 95% confidence limits for the pre-1989 rate and the post-1989 decay are shown by the thin grey lines. The decay obeys Omori's law (see text); the aftershock duration, $t_a = 5-28$ yr at 95% confidence. **c**, Seismicity rate drops on the southern Hayward fault, 40-80 km north of Loma Prieta, and then recovers during the following ~ 8 years. (Some 30 km southeast of the south Hayward fault, the 1984 $M_w = 6.2$ Morgan Hill shock disrupts the seismicity during 1984-88, so this period is not used to estimate the pre-1989 seismicity rate. The highest values exceed $20 \text{ shocks yr}^{-1}$ during this period.)

than 40 km from the Loma Prieta fault—well outside the traditional aftershock zone—the response to a stress change of either sign is a sudden seismicity rate change followed by a recovery roughly to the former rate³² (Fig. 6). The close resemblance between the observed behaviour of seismicity following such a stress increase (Fig. 6b) and that modelled with equation (2) (Fig. 3, inset) is evident.

Although, for small shocks, seismicity rate changes and probabilities can be tested against observations, the calculated probabilities for large earthquakes are more difficult to validate because there are so few large shocks. Probability calculations are also fraught with uncertainties associated with the prospective earthquake's location and magnitude, the variation of the earthquake inter-event time, and the probability density function (how the probability grows with time). Monte Carlo simulation can capture the range of behaviour, but this range can be frustratingly large. But the probability changes associated with earthquake transfer are less sensitive to these assumptions, enabling inferences to be made about how the probability of an earthquake on one fault is affected by a nearby large earthquake on another.

Several attempts to estimate earthquake probabilities using stress transfer and state and rate friction have been made. The seven $M_w \geq 6.8$ 'falling-domino' shocks on the North Anatolia fault during 1939-67 are obvious candidates, because it is difficult to explain such a rapid and progressive sequence unless transient probability gains associated with stress transfer are included. In 1997, my colleagues and I calculated that there was an average 3-fold probability gain at the site of each successive event caused by the preceding earthquakes³³. We also identified two segments—two dominoes still standing—where the stress rise since 1939 was greatest. We calculated a 15% 30-year probability for a large earthquake near Erzincan, a 12% probability near Izmit, but only a $< 1\%$ probability on the remaining 750 km of the fault. The 17 August

1999 $M_w = 7.4$ Izmit shock occurred in one of these stressed sites. The Izmit earthquake, in turn, increased the Coulomb stress and rate of seismicity on the Düzce fault, located east of the August rupture, and the Yalova fault, located west of the August rupture⁴⁶. The 12 November 1999 $M_w = 7.2$ Düzce earthquake subsequently struck on the Düzce fault where the stress was increased by ~ 3 bar.

Japan and California have proved to be important sites for testing such interaction-based probabilities. Toda *et al.*³⁴ calculated a 10-fold probability drop during the next 30 years on the section of the major fault most discouraged from failure by the 1995 $M_w = 6.9$ Kobe shock, and a 5-fold probability gain on the section most stressed, which lies near Kyoto. Three studies have focused on the San Francisco Bay area where, as in the other cases, the consequences of an urban $M_w = 7$ shock loom large, but the interaction of several sub-parallel faults is simpler than in Japan. Building on the stress analysis of Jaumé and Sykes³⁵, Harris and Simpson³⁶ retrospectively evaluated the suppression of $M_w \geq 6$ shocks in the San Francisco Bay area after 1906, finding the set of stress and state/rate constitutive parameters consistent with the observed rate change. Prospective studies of the probability of a $M_w = 6.8$ shock on the Hayward fault during 2000–30, such as occurred in 1868, find that the probability is 15–25% lower if the effect of the 1906 shock is included^{32,37}.

These nascent probability calculations are faithful to the presence of aftershocks, the characteristics of large earthquake sequences, the change in location and style of upper-plate earthquakes following subduction events³⁸, and the cessation of large shocks in the San Francisco Bay area after 1906. Perhaps most surprising, one finds that the stress trigger or shadow cast by a great earthquake can exert an influence on earthquake occurrence for more than a century. What lies ahead is fuller incorporation of viscoelastic^{39–41} and poroelastic effects that modify the Coulomb stress with time, real-time tests of earthquake rate and probability calculations, and further exploration of seismicity rate changes and the dynamic Coulomb stresses. □

1. Bakun, W. H. Seismic activity of the San Francisco Bay region. *Bull. Seismol. Soc. Am.* **89**, 764–784 (1999).
2. Lay, T., Astiz, L., Kanamori, H. & Christensen, D. H. Temporal variation of large interplate earthquakes in coupled subduction zones. *Phys. Earth Planet. Inter.* **54**, 258–312 (1989).
3. Chinnery, M. A. The stress changes that accompany strike-slip faulting. *Bull. Seismol. Soc. Am.* **53**, 921–932 (1963).
4. Das, S. & Scholz, C. H. Off-fault aftershock clusters caused by shear stress increase? *Bull. Seismol. Soc. Am.* **71**, 1669–1675 (1981).
5. Stein, R. S. & Lisowski, M. The 1979 Homestead Valley earthquake sequence, California: Control of aftershocks and postseismic deformation. *J. Geophys. Res.* **88**, 6477–6490 (1983).
6. Oppenheimer, D. H., Reasenber, P. A. & Simpson, R. W. Fault plane solutions for the 1984 Morgan Hill, California, earthquake sequence: Evidence for the state of stress on the Calaveras fault. *J. Geophys. Res.* **93**, 9007–9026 (1988).
7. King, G. C. P., Stein, R. S. & Lin, J. Static stress changes and the triggering of earthquakes. *Bull. Seismol. Soc. Am.* **84**, 935–953 (1994).
8. Stein, R. S., King, G. C. P. & Lin, J. Change in failure stress on the southern San Andreas fault system caused by the 1992 magnitude = 7.4 Landers earthquake. *Science* **258**, 1328–1332 (1992).
9. Harris, R. A. & Simpson, R. W. Changes in static stress on southern California faults after the 1992 Landers earthquake. *Nature* **360**, 251–254 (1992).
10. Jaumé, S. C. & Sykes, L. R. Change in the state of stress on the southern San Andreas fault resulting from the California earthquake sequence of April to June 1992. *Science* **258**, 1325–1328 (1992).
11. Gross, S. & Kisslinger, C. Estimating tectonic stress rate and state with Landers aftershocks. *J. Geophys. Res.* **102**, 7603–7612 (1997).
12. Harris, R. A. Introduction to special session: Stress triggers, stress shadows, and implications for seismic hazard. *J. Geophys. Res.* **103**, 24347–24358 (1998).
13. Reasenber, P. A. & Simpson, R. W. Response of regional seismicity to the static stress change produced by the Loma Prieta earthquake. *Science* **255**, 1687–1690 (1992).
14. Simpson, R. W. & Reasenber, P. A. in *The Loma Prieta, California, Earthquake of October 17, 1989—Tectonic Processes and Models* (ed. Simpson, R. W.) F55–F89 (Professional Paper 1550-F, US Geological Survey, 1994).
15. Parsons, T., Stein, R. S., Simpson, R. W. & Reasenber, P. A. Stress sensitivity of fault seismicity: A

- comparison between limited-offset oblique and major strike-slip faults. *J. Geophys. Res.* **104**, 20183–20202 (1999).
16. Zoback, M. D. *et al.* New evidence on the state of stress of the San Andreas fault system. *Science* **238**, 1105–1111 (1987).
17. Rice, J. R. in *Fault Mechanics and Transport Properties of Rock* (eds Evans, B. & Wong, T.-F.) 475–504 (Academic, London, 1992).
18. Cocco, M., Nostro, C. & Ekström, G. Static stress changes and fault interaction during the 1997 Umbria-Marche earthquake sequence. *J. Geophys. Res.* (in the press).
19. Seeber, L. & Armbruster, J. G. Earthquakes as beacons of stress change. *Nature* (submitted).
20. Hardebeck, J. L., Nazareth, J. J. & Hauksson, E. The static stress change triggering model: Constraints from two southern California aftershocks sequences. *J. Geophys. Res.* **103**, 24427–24437 (1998).
21. Anderson, G. & Johnson, H. A new statistical test for static stress triggering: Application to the 1987 Superstition Hills earthquake sequence. *J. Geophys. Res.* **104**, 20153–20168 (1999).
22. Belardinelli, M. E., Cocco, M., Coutant, O. & Cotton, F. Redistribution of dynamic stress during coseismic ruptures: Evidence for fault interaction and earthquake triggering. *J. Geophys. Res.* **104**, 14925–14946 (1999).
23. Kilb, D., Gombert, J. & Bodin, P. Aftershock triggering by dynamic and static Coulomb stress changes. *Eos (suppl.)* **79**, F647 (1998).
24. Hill, D. P. *et al.* Seismicity remotely triggered by the magnitude 7.3 Landers, California, earthquake. *Science* **260**, 1617–1623 (1993).
25. Vidale, J. E., Agnew, D. C., Johnston, M. S. J. & Oppenheimer, D. H. Absence of earthquake correlation with earth tides: An indication of high pre-seismic fault stress rate. *J. Geophys. Res.* **103** (1998).
26. Lockner, D. A. & Beeler, N. M. Premonitory slip and tidal triggering of earthquakes. *J. Geophys. Res.* **104**, 20133–20152 (1999).
27. Vidale, J., Agnew, D., Oppenheimer, D., Rodriguez, C. & Houston, H. A weak correlation between earthquakes and extensional normal stress and stress rate from lunar tides. *Eos (suppl.)* **79**, F641 (1998).
28. Working Group on the probabilities of future large earthquakes in southern California. *Future Seismic Hazards in Southern California, Phase I: Implications of the 1992 Landers Earthquake Sequence* (California Division of Mines & Geology, Sacramento, 1992).
29. Working Group on California earthquake probabilities. Seismic hazards in southern California: Probable earthquakes. 1994–2014. *Bull. Seismol. Soc. Am.* **85**, 379–439 (1995).
30. Dieterich, J. A constitutive law for rate of earthquake production and its application to earthquake clustering. *J. Geophys. Res.* **99**, 2601–2618 (1994).
31. Dieterich, J. H. & Kilgore, B. Implications of fault constitutive properties for earthquake prediction. *Proc. Natl. Acad. Sci. USA* **93**, 3787–3794 (1996).
32. Parsons, T. & Stein, R. Earthquake probabilities in the San Francisco Bay area after 1906: A rate/state approach to fault interactions. *J. Geophys. Res.* (submitted).
33. Stein, R. S., Barka, A. A. & Dieterich, J. H. Progressive failure on the North Anatolian fault since 1939 by earthquake stress triggering. *Geophys. J. Int.* **128**, 594–604 (1997).
34. Toda, S., Stein, R. S., Reasenber, P. A. & Dieterich, J. H. Stress transferred by the $M_w = 6.5$ Kobe, Japan, shock: Effect on aftershocks and future earthquake probabilities. *J. Geophys. Res.* **103**, 24543–24565 (1998).
35. Jaumé, S. C. & Sykes, L. R. Evolution of moderate seismicity in the San Francisco Bay region, 1850 to 1993: Seismicity changes related to the occurrence of large and great earthquakes. *J. Geophys. Res.* **101**, 765–789 (1996).
36. Harris, R. A. & Simpson, R. W. Suppression of large earthquakes by stress shadows: A comparison of Coulomb and rate-and-state. *J. Geophys. Res.* **103**, 24439–24451 (1998).
37. Reasenber, P. A., Simpson, R. W., Matthews, M. V., Ellsworth, W. L. & Parsons, T. Recurrence models and fault interactions: an example based on the 1868 Hayward, California, earthquake. *Bull. Seismol. Soc. Am.* (submitted).
38. Taylor, M. A., Dmowska, R. & Rice, J. R. Upper plate stressing and seismicity in the subduction earthquake cycle. *J. Geophys. Res.* **103**, 24523–24542 (1998).
39. Pollitz, F. F. & Sacks, I. S. The 1995 Kobe, Japan, earthquake: A long-delayed aftershock of the offshore 1944 Tonankai and 1946 Nankaido earthquakes. *Bull. Seismol. Soc. Am.* **87**, 1–10 (1997).
40. Freed, A. M. & Lin, J. Time-dependent changes in failure stress following thrust earthquakes. *J. Geophys. Res.* **103**, 24393–24410 (1998).
41. Kenner, S. & Segall, P. Time-dependence of the stress shadowing effect and its relation to the structure of the lower crust. *Geology* **27**, 119–122 (1999).
42. Bakun, W. H. Scenarios for historic San Francisco Bay region earthquakes. *US Geol. Surv. Open File Rep.* **98–785** (1998).
43. Stein, R. S., King, G. C. P. & Lin, J. Stress triggering of the 1994 $M = 6.7$ Northridge, California, earthquake by its predecessors. *Science* **265**, 1432–1435 (1994).
44. Reasenber, P. A. Seismicity patterns in southern California before and after the 1994 Northridge earthquake: A preliminary report. *US Geol. Surv. Open-File Rep.* **95–484** (1995).
45. Hardebeck, J. L. & Hauksson, E. Background Stress State Plays a Role in Earthquake Triggering. *Eos (suppl.)* (in the press).
46. Barka, A. The 17 August 1999 Izmit earthquake. *Science* **285**, 1858–1859 (1999).

Acknowledgements

I thank the many colleagues who shared their preliminary research, and R. Dmowska, J. Lin, R. Madariaga, T. Parsons, F. Pollitz, J. Rice and S. Toda for comments. This Review Article was written while at Ecole Normale Supérieure and the Institut de Physique du Globe de Paris. The support of Pacific Gas & Electric Co. is gratefully acknowledged.

Correspondence and requests for materials should be addressed to the author (e-mail: rstein@usgs.gov).

Distribution of spatial and nonspatial information in dorsal hippocampus

Robert E. Hampson, John D. Simeral & Sam A. Deadwyler

Department of Physiology & Pharmacology, Neuroscience Program, Wake Forest University School of Medicine, Winston-Salem, North Carolina 27157, USA

The hippocampus in the mammalian brain is required for the encoding of current and the retention of past experience. Previous studies have shown that the hippocampus contains neurons that encode information required to perform spatial and nonspatial short-term memory tasks. A more detailed understanding of the functional anatomy of the hippocampus would provide important insight into how such encoding occurs. Here we show that hippocampal neurons in the rat are distributed anatomically in distinct segments along the length of the hippocampus. Each longitudinal segment contains clusters of neurons that become active when the animal performs a task with spatial attributes. Within these same segments are ordered arrangements of neurons that encode the nonspatial aspects of the task appropriate to those spatial features. Thus, anatomical segregation of spatial information, together with the interleaved representation of nonspatial information, represents a structural framework that may help to resolve conflicting views of hippocampal function.

Current theories of hippocampal function wrestle with two major conceptual issues: whether the structure encodes spatial, nonspatial, or both types of information^{1–7} and whether the information encoded is episodic or semantic^{8–11}. At the heart of this controversy is a lack of agreement as to what is represented by the firing of hippocampal cells¹²: stimuli and behavioural events in the form of ‘memories’^{13,14}, or information that is exclusively used for spatial navigation^{15,16}. Investigators have begun to examine this issue by determining how information is distributed across the septotemporal axis of the hippocampus^{17–19}.

The hippocampus receives multimodal sensory information from many neocortical structures; however, unlike primary sensory and associational areas of the brain, it does not appear to represent that information in any obvious topographic format²⁰. However, lesion studies in the rat have indicated that behaviours requiring spatial as opposed to nonspatial information are differentially affected depending on where damage occurs along the longitudinal axis of the hippocampus²¹. Investigations of hippocampal cell firing have shown robust cross-correlations between ‘place cells’ with overlapping spatial areas, even when those cells are separated by large distances within the hippocampus²². However, the lack of a functionally relevant structure within the hippocampus to show how encoded information is distributed across neurons^{23,24} is an important puzzle concerning hippocampal processing. We provide evidence here of a unique anatomical framework for representation of spatial and nonspatial information in the hippocampus, based on overlapping distributions of cells encoding task-related events.

Distributed hippocampal ensembles encode events

Multi-electrode recording of ensembles of hippocampal neurons was conducted in 23 rats performing a spatial delayed-nonmatching-to-sample (DNMS) short-term memory task (Fig. 1a, see Methods). The DNMS task is hippocampus-dependent, in that selective removal of the hippocampus disrupts²⁵, and glutamatergic receptor modulation enhances²⁶, performance. We determined how information was encoded in the hippocampus during the task by simultaneously recording from ensembles of 10–16 individual neurons using specially constructed microwire electrode arrays and spike-sorting techniques (Fig. 1b). Eight pairs of recording electrodes were positioned along the septotemporal axis of the dorsal hippocampus at 200- μ m intervals over a total expanse of 1.6 mm. One electrode of each pair recorded from area CA1, the other from area CA3. Single cells were isolated at each electrode location using extracellular action potential criteria^{14,27}. All cells

reported were recorded continuously throughout at least five DNMS sessions in well trained animals (see Methods).

Population analysis of DNMS trial-specific neural ensemble activity is sufficient to predict both correct and incorrect performance^{27,28} (Fig. 1c). Figure 1d shows examples of hippocampal neurons recorded under these circumstances, which demonstrated several firing patterns unique to DNMS task requirements. ‘Position’ cells increased firing differentially to a response on either the left or the right lever, irrespective of when in the trial the response occurred (Fig. 1a). ‘Phase’ cells fired during the response on the lever in only one phase (for example, sample) of the task but not the other (for example, nonmatch), and firing was irrespective of lever position. Other neurons were further differentiated with respect to firing correlates and could be classified in terms of combinations or ‘conjunctions’ of the above two features (for example, right-sample or left-nonmatch); these were called ‘conjunctive’ cells^{13,14,24} (Fig. 1d). Other cells recorded during the DNMS task fired specifically in response to the two (out of four) possible events that made up a type of trial (for example, a right-sample response and a left-nonmatch response). These were designated as ‘trial-type’ cells, as they encoded only one of the two types of trial (Fig. 1d).

Anatomical distribution of cell firing correlates

On the basis of the above firing repertoire, 243 cells were classified into the above four categories across 23 different animals. As all neurons in the above categories also had fixed electrode locations on the array which indicated relative anatomical position along the septotemporal axis of the hippocampus (Fig. 1b), the correspondence between location and cell type could be determined. We constructed composite maps of cell locations across all animals by plotting the locations of cells (identified with respect to the above four types) on fold-out maps of hippocampus²². Cells with different task-relevant correlates were found to be organized in a spatiotopic manner within the hippocampus. Figure 2a shows the distribution of 214 cells (position, phase and conjunctive) plotted on a fold-out map of the CA3 and CA1 subregions over the 1.6–2.0-mm expanse covered by the electrode array in dorsal hippocampus (trial-type cells are not shown). There was no difference ($F_{(5,203)} = 1.71$, $P = 0.13$) between subregions. In Fig. 2b, c, the anatomic locations of position and phase cells (Fig. 1d) are plotted separately. Position cells were the fewest in number ($n = 33$), but were distributed within the largest longitudinal segments of hippocampus. Position cells for ‘left response’ were contained in 0.6–0.8-mm segments

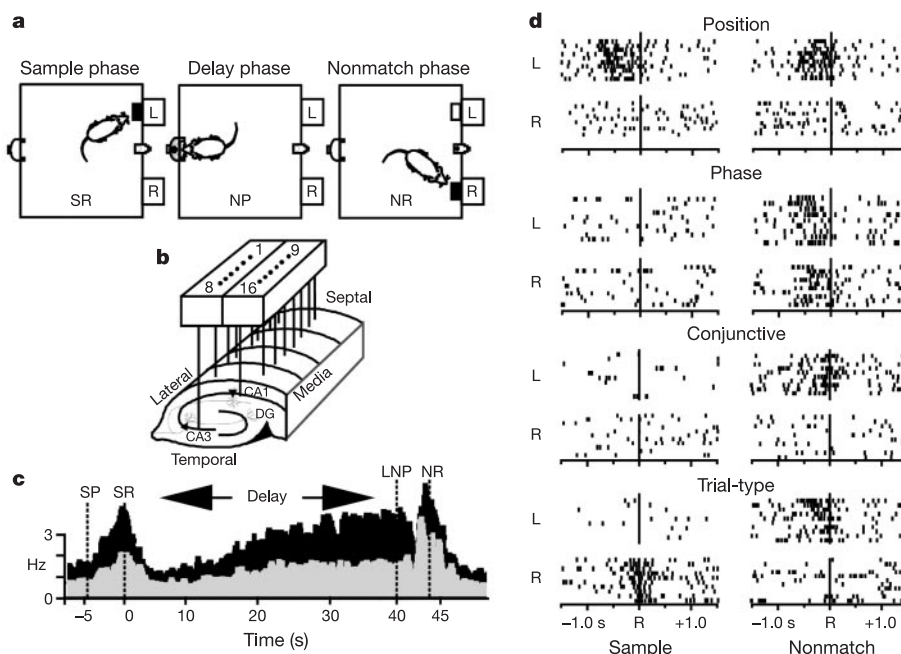


Figure 1 Multi-electrode recording during delayed-nonmatch-to-sample (DNMS) task. **a**, DNMS trial showing sample, delay and nonmatch phases of a left-sample, right-nonmatch trial. L, left lever position; R, right lever position; SR, sample response; NP, nosepoke during delay interval; NR, nonmatch response. **b**, Array of 16 microwires used for simultaneous recording of identified CA3 and CA1 neurons at indicated locations along the septotemporal axis of the hippocampus. **c**, Trial-based ensemble histograms. Black

histogram indicates firing on correct DNMS trials; grey histogram denotes firing on error trials. SP, sample presentation; LNP, last nosepoke in delay. **d**, Peri-event rastergrams from 10 individual trials (rows) of 4 hippocampal cell types (Position, Phase, Conjunctive and Trial-type) construct at ± 1.5 s around the response (R). The increased density of dots (extracellular spikes) at R indicates increased firing rates during the lever press response.

encompassing 3–4 electrode positions (blue squares, Fig. 2b), but were completely segregated from position cells which encoded ‘right response’ in the next (0.6–0.8 mm) longitudinal segment (red squares, Fig. 2b). The possibility that this number of cells would be distributed in this arrangement by chance was examined using an $R \times C$ goodness of fit test for independence²⁹ and rejected ($G_{(19)} = 83.6$, $P < 0.001$). The distribution of position cells was not significantly different ($G_{(19)} = 25.4$, $P = 0.17$) in CA3 and CA1. Recurrences of clusters of left and right position cells on either side of the two major (0.6–0.8 mm) distributions were present (see asterisks) in sectors 1, 9 and 10 of the fold-out map in Fig. 2b, indicating that position cell clusters may recur every 1.2–1.6 mm along the septotemporal axis.

A similar analysis of phase cells, which encoded which phase of the DNMS task the animal was in (sample or nonmatch), showed more frequently recurring clusters, but these were dispersed within the same septotemporal regions as position cell clusters. Phase cells specific for sample ($n = 38$) or nonmatch ($n = 42$) were distributed in alternating 0.2–0.4-mm segments in both the CA3 and CA1 subfields, interleaved with position cell clusters.

Phase cells appeared to be in register with position cell clusters in both CA1 and CA3, in that each 0.6–0.8-mm position cell segment contained clusters of both sample and nonmatch (phase) cells. This relationship is shown in Fig. 2c (left) by the red/blue (position) and green/yellow (phase) ribbon gradients. Clusters of phase cells also alternated in accordance with the fragmented repeats for position cells (Fig. 2, sectors 1, 9 and 10, asterisks). Thus, independent clusters of phase cells were systematically interleaved within the same anatomical regions of hippocampus that were occupied by position cells. Figure 2c shows a significant difference between sample and nonmatch phase cell locations in CA3 versus CA1 ($G_{(19)} = 49.9$, $P < 0.001$), reflecting the fact that the two subregions were out of register with respect to the positioning of these functional cell types. A similar ‘offset’ has been reported for cross-correlations between CA1 and CA3-located place cells²².

Distribution of conjunctive and trial-type cells

The interleaved distributions of position and phase cells (Fig. 2c) indicated that critical DNMS task information may be partitioned along the hippocampus in an anatomically specific manner. Given the anatomical constraints imposed by the distributions of the position and phase cells, localization of conjunctive cells (Fig. 1d) by the same method must either confirm or reject this organizational structure. Therefore, a given conjunctive cell that uniquely encodes ‘left-sample’ should, by definition, be located within the designated intersection of the appropriate position and phase cell regions of the hippocampus. Figure 3 shows the distributions of conjunctive cells ($n = 101$) superimposed on red/blue and green/yellow gradients which indicate the locations of position and phase cell clusters in Fig. 2c. Independent assessment of the positions of the four different subtypes of conjunctive cells showed that their distributions were nonrandom ($G_{(17)} = 87.2$, $P < 0.001$), and coincided precisely with the locations predicted by the intersection of the appropriate position and phase cell clusters (Figs 2 and 3). In addition, Fig. 3b–d shows fragmented repeats in the conjunctive cell distributions that correspond to the locations of repeats in the position cell gradients (Fig. 2b, asterisks). Finally, the ‘offset’ in phase cell gradients between CA3 and CA1 was duplicated in the distribution of conjunctive cells (Fig. 3b, c). The likelihood that this independently recorded population of conjunctive cells would be distributed with this degree of correspondence to the position and phase cell clusters by chance was, as expected, extremely low ($G_{(57)} = 468.9$, $P < 1 \times 10^{-9}$).

A third factor that supported the above anatomical arrangement of cell firing correlates was the location of 29 recorded trial-type cells (Fig. 1d). As such cells fire during both responses (sample and nonmatch) within a particular type of DNMS trial, they encode a combination of two different lever positions in the two different phases of the task. Exactly how these cells would be distributed with respect to the above anatomical arrangement was not obvious. However, the white stars in Fig. 3 show that trial-type cells tended

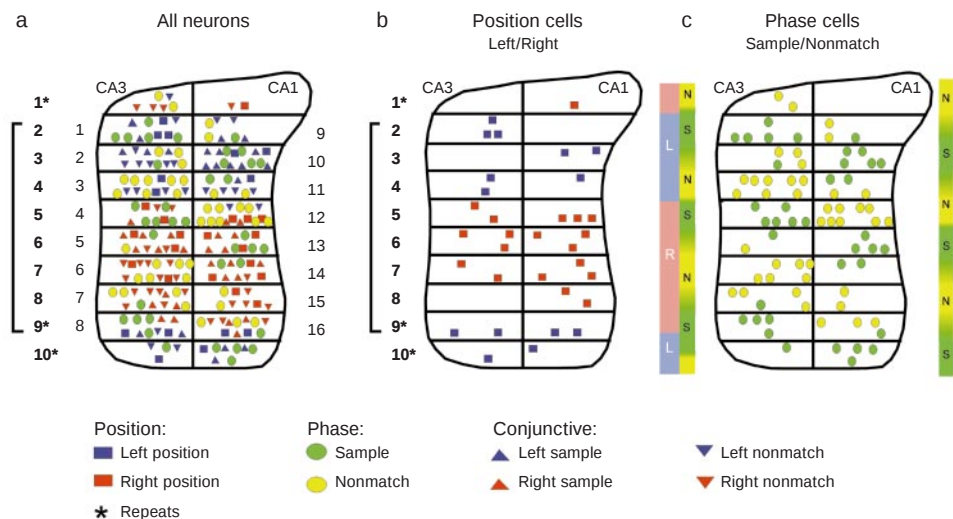


Figure 2 Anatomical localization of cell firing correlates. **a**, Distribution of 214 categorized hippocampal cells (trial-type cells excluded) plotted on ‘fold-out’ map of dorsal hippocampal CA3 and CA1 subfields. Electrode positions (Fig. 1b) are indicated by bracket and italic numbers at left and right as successive 200- μ m segments. Two additional positions (bold numbers) were revealed in composite maps across animals. **b**, Distribution of position cells shown in Fig. 1d for right and left responses. Sectors designated 1, 9 and

10 in the fold-out map (asterisks) indicate fragmented ‘repeats’ of position cell distributions. **c**, Distribution of phase cells (Fig. 1d) for sample and nonmatch phases of the task. Red/blue and green/yellow gradients indicate interleaved locations of phase (S,N) with position (L,R) cell clusters. Gradient at right illustrates ‘offset’ of phase cell clusters in CA1 versus CA3 subregions.

to be distributed near to or across the borders of position cell segments. At least one trial-type cell was present near each of the four different position cell borders in both CA1 and CA3, as well as near the repeats (Fig. 2b). Finally, trial-type cells, like conjunctive cells, were also offset in CA1 relative to CA3.

Temporal separation of task-relevant events

Figure 3 shows the two different types of DNMS trial as they would appear with respect to hippocampal cell activity during performance of the task. It is clear that this anatomical segregation maximizes the encoding of task-relevant features (position and phase cells) as well as trial-specific events (conjunctive and trial-type cells) within the hippocampus by ensuring that the clusters that encode incompatible features do not overlap anatomically. However, were it not for an important complement to this encoding strategy, the distinctiveness of the above location codes would be much less precise. Although there is no overlap between conjunctive cell clusters within a trial (Fig. 3a, b), there is only minimal anatomical separation (0.2–0.4 mm) between clusters of conjunctive cells encoding the same response position for different task phases (for example, left-sample and left-nonmatch; Fig. 3b, c) and trial types. However, because of the delay interval between the two phases (Fig. 1a), the same response position information within the ensemble never coexists with different phases of the DNMS trial (except in the case of an error)²⁸. Thus, conjunctive cells located in the right-sample region (Fig. 3a) of the hippocampus will tend not to fire at the same time as cells in the adjacent nonmatch regions (Fig. 3d), owing to temporal separation by the delay interval. This ensures that the position-specific segments of hippocampus will be activated differentially, temporally as well as spatially (that is, anatomically)³⁰, during performance of the task.

Functional significance of anatomical segregation

Our study complements the findings of ref. 31, which showed that hippocampal neurons encode spatial and nonspatial aspects of a continuous delayed match/nonmatch-to-sample odour-memory task. A similar, proportioned distribution of functional cell types was obtained within the population of hippocampal neurons and included similar parameters to those reported here (cup position, phase of task and so on) as well as combinations (conjunctions) of

those features. Our findings extend this global knowledge characteristic of hippocampal encoding to an anatomical framework in which location within the hippocampus determines which of those cells may be activated.

Given the functional anatomy depicted here, selective lesions of the dorsal hippocampus would impair performance as a direct consequence of which septotemporal level was affected. Evidence from selective lesion studies indicates that task-specific impairment results from damage to different septotemporal regions of hippocampus^{21,32}. From such findings it has been proposed that, normally, a widespread region of dorsal hippocampus (70%) is involved in the performance of spatial tasks²¹, but if animals are trained with as little as 25–30% of dorsal hippocampus intact, neither acquisition nor retention of the task is impaired. Thus, even a relatively small lesion targeted to the region of a particular position cell cluster (Fig. 2b) could severely impair spatial navigation, but only mildly impair tasks with information encoded in redundant fashion along the septotemporal axis (such as DNMS)²⁵.

The anatomical arrangement shown in Fig. 3 indicates that position cells would be activated only in appropriate regions of hippocampus consistent with the defined response position, but phase cells would fire in all spatial locations irrespective of which position cells were activated (Fig. 2c). However, as phase cells do not encode position, this does not violate the functional nature of the anatomical code. Phase cells will therefore still fire in concert with clusters of conjunctive and trial-type cells located within the anatomical regions defined by the activated position cells (Figs 2, 3). Taking this into account, a remarkably consistent $\sim 40\%$ of total cells ($n = 243$) per event could be classified as active within a given anatomical segment during any one of the four task-relevant events (LS, 96 cells, 39.5%; RS, 99 cells, 40.7%; LN, 99 cells, 40.7%; RN, 91 cells, 37.4%). As the four cell types encoded all information relevant to the task (Fig. 2a), their location and distribution within the hippocampus indicate that the demonstrated anatomical arrangement provides a basis for functional differentiation. Conjunctive cells could receive appropriate synaptic connections from phase and position cells located within the same anatomical region to produce their event-specific firing properties (for example, right-sample, Fig. 3a). Extending this rubric to trial-type cells implies that they could be the product of converging synaptic connections from

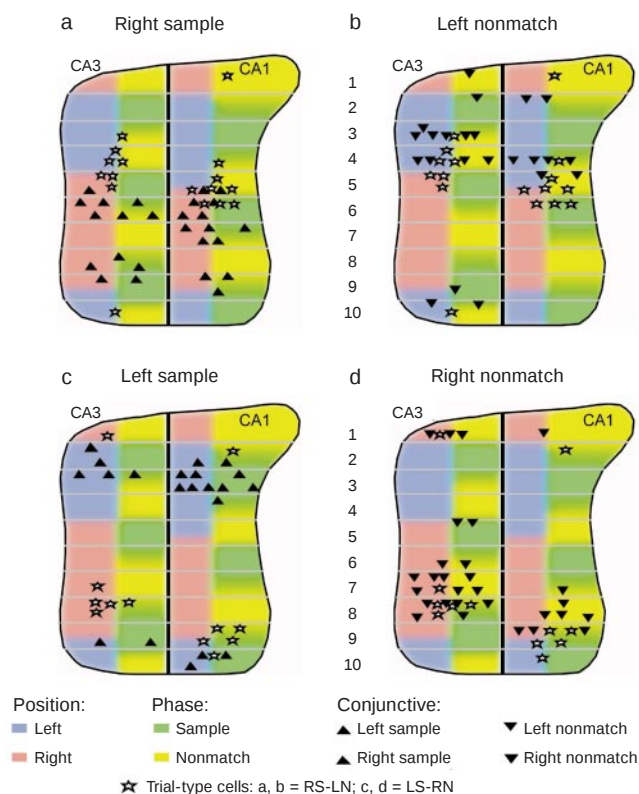


Figure 3 Distribution of conjunctive and trial-type cells mapped onto position and phase cell gradients shown in Fig. 2c. **a**, Right-sample conjunctive cells were found at the intersection of right position and sample phase gradients in both CA3 and CA1. **b**, Left-nonmatch conjunctive cells were located at the intersection of left position and nonmatch phase gradients. Stars in **a** and **b** depict locations of trial-type cells (Fig. 1d) which fired during both phases (RS-LN) of the trial. **c**, Conjunctive cells for left-sample phase of the task. Phase and position gradients as in **a** and **b**. **d**, Conjunctive cells for right-nonmatch phase of the task. Stars: LS-RN trial-type cells. Note different locations of position and phase gradients. CA3 misalignment in **c** was the only violation of the anatomical scheme.

different conjunctive cells (for example, a right-sample and a left-nonmatch). Such hierarchical arrangements are prevalent in other brain regions where categorizations of stimulus components and conjunctions of stimulus elements are encoded for other purposes^{33,34}.

With the exception of the 0.2–0.4-mm offset, both the information encoded by CA1 cells and their relative septotemporal distribution were virtually identical to those of CA3 cells. This suggests either that the demonstrated functional attributes (Fig. 1d) were imposed on both subregions independently by similar extrinsic inputs, or that the functional properties of CA1 cells were in fact generated by connections with CA3 neurons. Connections between CA3 and CA1 neurons in the rat have been reported to have sufficient synaptic interactions to generate the observed distributions of cell types within these two subregions^{35–37}. The 0.2–0.4-mm offset of the entire cluster arrangement in CA1 relative to CA3 could reflect functional Schaeffer collateral projections in providing ‘lamellar-like’³⁸ connections between the two subregions. Recent anatomical evidence indicates that axonal projections from the entorhinal cortex to the dentate gyrus have extensive longitudinal spread along the septotemporal axis of the hippocampus³⁹. However, the projections from the dentate hilus to other regions of the hippocampus (CA3) are highly segregated within these larger longitudinal entorhinal projection zones³². The anatomical organization shown in Fig. 3 is compatible with such a subdivision of

larger areas of hippocampus (position cells) into a series of overlapping discrete projections to more restricted cell clusters (conjunctive cells). The portioning of different groups of functional cell types along the longitudinal axis of hippocampus provides a basis for cell selection by the major intrinsic and extrinsic pathways, whose trajectories are oriented in an oblique manner to the septotemporal axis³⁷.

The CA3 pyramidal cell region of hippocampus has been described as an autoassociative network^{6,40}. The ‘binding’ of information by hippocampal neurons has been addressed in animal studies of the encoding of spatial as well as nonspatial attributes^{41,42}, and in human studies using functional imaging techniques^{8,11,18,19}. Such networks are ubiquitous in many brain regions, but they share the common property of self-organization, where selected cells within the network respond to distributed inputs in a ‘winner-take-all’ fashion⁴³. The arrangement presented here, showing rigorous segregation and interleaving of dimensional (position and phase cells) and more precisely encoded trial-specific features (conjunctive and trial-type cells), is suggestive of an autoassociative network. The formation of the observed clusters could occur through repetitive patterns of specific activity that ‘sculpt out’ neuronal firing within ensembles in response to the recurrence of combinations of behaviourally significant events. If the neurons responsible for such patterns were located within discrete anatomical regions, the clustering of functionally distinct cell types would be readily detected by the multi-neuron recording methods employed here. In addition, the nature of the distributed ensemble firing pattern would correspond closely to the proposed hippocampal topography shown in Fig. 3.

The categorization of functional cell types and their selective location along the septotemporal axis of the hippocampus topographically represent both spatial and nonspatial information in a task-relevant manner. The slight prioritization of the anatomical arrangement for spatial (position) as opposed to nonspatial (phase) elements of the task may reflect the propensity to encode spatial attributes as a ‘default’ mode of operation when other co-features (conjunctions) are not present or have not been established through behavioural or cognitive manipulations⁴⁴. In this regard, it is not difficult to imagine how cells encoding only spatial attributes (position cells) could be activated on some occasions^{30,45}, whereas on other occasions nonspatial attributes would also be detected^{13,46}. Perhaps the most important aspect of these findings is that both types of information can be encoded within the hippocampus in a manner that is consistent with the mnemonic demands of the task⁴⁷. □

Methods

Behavioural training

Male Long-Evans rats ($n = 23$) were trained to criterion on a two-lever spatial delayed-nonmatching-to-sample (DNMS) task with randomly occurring variable delay intervals of 1–40 s (ref. 14). Animals performed the task by pressing (sample response, SR) a single lever presented (sample presentation, SP) in one of the two positions in the sample phase (left or right). The lever was then retracted and the delay phase initiated, for the duration of which the animal was required to nosepoke into a lighted device on the opposite wall (Fig. 1a). Following termination of the delay the nosepoke light was extinguished, both levers were extended and the animal was required to press the lever opposite to the sample lever (nonmatch response, NR). If the correct lever was pressed the animal was rewarded with a drop of water (Fig. 1a) and the trial was completed. The longer the delay between sample and nonmatch phases (1–40 s), the less likely the animal was to be correct on that trial^{42,2}. A 10-s intertrial interval (ITI) was imposed between all trials. Incorrect responses produced a darkened chamber for 5 s without reward, followed by the final 5 s of the ITI. Sessions consisted of 100–150 consecutive trials, administered five days per week. Only data from correct trials are included in the analyses.

Multi-neuron recording

Arrays of 16 microwires (40 μm) were surgically implanted using both stereotactic coordinates and spontaneous cell firing activity to position electrode tips in the CA3 and CA1 cell layers (Fig. 1b)⁴⁸. All arrays were supplied by NB Labs and had a fixed geometry with 200 μm between pairs (septotemporal axis) and 800 μm between rows (medial–lateral axis). Data were collected from 23 animals and a total of 243 cells (average 11.5 per

animal, range 9–16) over at least five DNMS sessions with stable extracellular action potential waveforms and consistent event-specific firing patterns (Fig. 1d).

Data analysis

Hippocampal cell types associated with specific behavioural events on each trial were sorted into categories of behavioural events as described in the text. Peri-event histograms (± 1.5 s, 100-ms bins) were constructed around each event for each neuron. Peak firing frequency (max) within ± 1.0 s of the event relative to baseline firing (± 1.5 s to ± 1.0 s) before the event was transformed into a z-score (max – baseline)/s.d. of baseline). The neuron was considered to encode the event if its firing rate exceeded a z-score of 3.19 ($P < 0.001$). To map cell locations onto composite CA3/CA1 fold-out maps, we plotted electrode locations from all arrays and evaluated them using a numerical transformation in which each individual animal's array contributed algebraically to a composite overall template of cell distribution. There was a highly significant correlation of the location of cells on each individual animal's array with the overall template, even without realignment ($\chi^2_{(2)} = 47.3$, $P < 0.001$, correlation $r^2 = 0.63$, chance $r^2 = 0.38$), attesting to the precision of stereotactic placement of arrays across animals. To optimize this pattern, the arrays of individual animals were realigned by 'sliding' the entire array in ± 200 μ m (one electrode pair) increments relative to the overall template. The overall template was then averaged again and the realignment of individual arrays with the template repeated. This was performed until maximal correlation with the overall template was achieved. Out of 23 cases it was only necessary to slide particular arrays a single electrode position (± 200 μ m) to optimize the composite maps shown in Figs 2 and 3, which improved representation of the cell clusters by 10% ($\chi^2_{(2)} = 72.9$, $r^2 = 0.73$, $P < 0.001$). Moving individual animal arrays more than one electrode position relative to the overall template decreased the correlation significantly and tended to dissolve the cell clusters. The likelihood that the overall pattern of cell types could be obtained strictly from the combination and alignment procedures was tested by 'shuffling' the cell types across the different electrode positions. The shuffling procedures did not produce coherent patterns across animals. Statistical comparisons of cell locations are reported using an $R \times C$ goodness of fit test (G-test) with Chi-square distribution and Williams adjustment for low n (ref. 29).

Received 27 August; accepted 8 October 1999.

- O'Keefe, J. A. & Nadel, L. *The Hippocampus as a Cognitive Map* (Clarendon, Oxford, 1978).
- Olton, D. S., Becker, J. T. & Handelmann, G. E. Hippocampus, space and memory. *Behav. Brain Sci.* **2**, 313–352 (1979).
- Rawlins, J. N. P. & Olton, D. S. The septo-hippocampal system and cognitive mapping. *Behav. Brain Res.* **5**, 331–358 (1982).
- Nadel, L. The hippocampus and space revisited. *Hippocampus* **1**, 221–229 (1991).
- Cohen, N. J. & Eichenbaum, H. The theory that wouldn't die: a critical look at the spatial mapping theory of hippocampal function. *Hippocampus* **1**, 265–268 (1991).
- Morris, R. G. & Frey, U. Hippocampal synaptic plasticity: role in spatial learning or the automatic recording of attended experience? *Phil. Trans. R. Soc. Lond. B. Biol. Sci.* **29**, 1489–1503 (1997).
- Muller, R. U., Stead, M. & Pach, J. The hippocampus as a cognitive graph. *J. Gen. Physiol.* **107**, 663–694 (1996).
- Tulving, E. & Markowitsch, H. J. Memory beyond the hippocampus. *Curr. Opin. Neurobiol.* **7**, 209–216 (1997).
- Vargha-Khadem, F. *et al.* Differential effects of early hippocampal pathology on episodic and semantic memory. *Science* **277**, 376–380 (1997).
- Squire, L. R. & Zola, S. M. Episodic memory, semantic memory, and amnesia. *Hippocampus* **8**, 205–211 (1998).
- Cohen, N. J. *et al.* Hippocampal system and declarative (relational) memory: Summarizing the data from functional neuroimaging studies. *Hippocampus* **9**, 83–98 (1999).
- Muller, R. A quarter of a century of place cells. *Neuron* **17**, 813–822 (1996).
- Shapiro, M. L., Tanila, H. & Eichenbaum, H. Cues that hippocampal place cells encode: dynamic and hierarchical representation of local and distal stimuli. *Hippocampus* **7**, 624–642 (1997).
- Deadwyler, S. A., Bunn, T. & Hampson, R. E. Hippocampal ensemble activity during spatial delayed-nonmatch-to-sample performance in rats. *J. Neurosci.* **16**, 354–372 (1996).
- Touretzky, D. S. & Redish, A. D. Theory of rodent navigation based on interacting representations of space. *Hippocampus* **6**, 247–270 (1996).
- Samsonovich, A. & McNaughton, B. L. Path integration and cognitive mapping in a continuous attractor neural network model. *J. Neurosci.* **17**, 5900–5920 (1997).
- Lepage, M., Habib, R. & Tulving, E. Hippocampal PET activations of memory encoding and retrieval: The HIPER model. *Hippocampus* **8**, 313–322 (1998).
- Gabrieli, J. D., Brewer, J. B. & Poldrack, R. A. Images of medial temporal lobe functions in human learning and memory. *Neurobiol. Learn. Mem.* **70**, 275–283 (1998).
- Schacter, D. L. & Wagner, A. D. Medial temporal lobe activation in fMRI and PET studies of episodic encoding and retrieval. *Hippocampus* **9**, 7–24 (1999).

- Burwell, R. D., Wittner, M. P. & Amaral, D. G. Perirhinal and postrhinal cortices of the rat: a review of the neuroanatomical literature and comparison with findings from the monkey brain. *Hippocampus* **5**, 390–408 (1995).
- Moser, M. B. & Moser, E. I. Distributed encoding and retrieval of spatial memory in the hippocampus. *J. Neurosci.* **18**, 7535–7542 (1998).
- Hampson, R. E., Byrd, D. R., Konstantopoulos, J. K., Bunn, T. & Deadwyler, S. A. Hippocampal place fields and spike-train correlation: Relationship between degree of place field overlap and cross-correlation. *Hippocampus* **6**, 281–293 (1996).
- Eichenbaum, H., Wiener, S. I., Shapiro, M. L. & Cohen, N. J. The organization of spatial coding in the hippocampus: A study of neural ensemble activity. *J. Neurosci.* **9**, 2764–2775 (1989).
- Eichenbaum, H. Thinking about cell assemblies. *Science* **261**, 993–994 (1993).
- Hampson, R. E., Jarrard, L. E. & Deadwyler, S. A. Effects of ibotenate hippocampal destruction on delayed matching and nonmatching to sample behavior in rats. *J. Neurosci.* **19**, 1492–1507 (1999).
- Hampson, R. E., Rogers, G., Lynch, G. & Deadwyler, S. A. Facilitative effects of the ampakine CX516 on short-term memory in rats: Enhancement of delayed-nonmatch-to-sample performance. *J. Neurosci.* **18**, 2740–2747 (1998).
- Hampson, R. E. & Deadwyler, S. A. Ensemble codes involving hippocampal neurons are at risk during delayed performance tests. *Proc. Natl Acad. Sci. USA* **93**, 13487–13493 (1996).
- Hampson, R. E., Rogers, G., Lynch, G. & Deadwyler, S. A. Facilitative effects of the ampakine CX516 on short-term memory in rats: Correlations with hippocampal ensemble activity. *J. Neurosci.* **18**, 2748–2763 (1998).
- Pedhazur, E. J. & Schmelkin, L. P. *Measurement, Design and Analysis: An Integrated Approach* (Erlbaum, Hillsdale, New Jersey, 1991).
- Colombo, M., Fernandez, T., Nakamura, K. & Gross, C. G. Functional differentiation along the anterior-posterior axis of the hippocampus in monkeys. *J. Neurophysiol.* **80**, 1002–1005 (1998).
- Wood, E., Dudchenko, P. A. & Eichenbaum, H. The global record of memory in hippocampal neuronal activity. *Nature* **397**, 613–616 (1999).
- Moser, M. B. & Moser, E. I. Functional differentiation in the hippocampus. *Hippocampus* **8**, 608–619 (1998).
- Hirsch, J. A., Alonso, J. M., Reid, R. C. & Martinez, L. M. Synaptic integration in striate cortical simple cells. *J. Neurosci.* **18**, 9517–9528 (1998).
- Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol. (Lond.)* **160**, 106–154 (1962).
- Ishizuka, N., Weber, J. & Amaral, D. G. Organization of intrahippocampal projections originating from CA3 pyramidal cells in the rat. *J. Comp. Neurol.* **295**, 580–623 (1990).
- Li, X. G., Somogyi, P., Ylinen, A. & Buzsaki, G. The hippocampal CA3 network: an *in vivo* intracellular labeling study. *J. Comp. Neurol.* **339**, 181–208 (1994).
- Amaral, D. G. & Witter, M. P. The three-dimensional organization of the hippocampal formation: a review of anatomical data. *Neuroscience* **31**, 571–591 (1989).
- Andersen, P., Bliss, T. V. P. & Skrede, K. K. Lamellar organization of hippocampal excitatory pathways. *Exp. Brain Res.* **13**, 222–238 (1971).
- Dolorfo, C. L. & Amaral, D. G. Entorhinal cortex of the rat: topographic organization of the cells of origin of the perforant path projection to the dentate gyrus. *J. Comp. Neurol.* **398**, 25–48 (1998).
- Wallenstein, G. V., Eichenbaum, H. & Hasselmo, M. E. The hippocampus as an associator of discontinuous events. *Trends Neurosci.* **21**, 317–323 (1998).
- Gothard, K. M., Skaggs, W. E., Moore, K. M. & McNaughton, B. L. Binding of hippocampal CA1 neural activity to multiple reference frames in a landmark-based navigation task. *J. Neurosci.* **16**, 823–835 (1996).
- McClelland, J. L. Complementary learning systems in the brain. A connectionist approach to explicit and implicit cognition and memory. *Ann. NY Acad. Sci.* **843**, 153–169 (1998).
- Kohonen, T. & Hari, R. Where the abstract feature maps of the brain might come from. *Trends Neurosci.* **22**, 135–139 (1999).
- Stern, C. E. *et al.* The hippocampal formation participates in novel picture encoding: evidence from functional magnetic resonance imaging. *Proc. Natl Acad. Sci. USA* **93**, 8660–8665 (1996).
- Burgess, N. & O'Keefe, J. Neuronal computations underlying the firing of place cells and their role in navigation. *Hippocampus* **6**, 749–762 (1996).
- Rawlins, J. N. P. A place for space and smells. *Nature* **397**, 561–563 (1999).
- Deadwyler, S. A. & Hampson, R. E. Ensemble activity and behavior: What's the code? *Science* **270**, 1316–1318 (1995).
- Deadwyler, S. A. & Hampson, R. E. *Annu. Rev. Neurosci.* **20**, 217–244 (1997).

Acknowledgements

We thank D. Byrd, J. Konstantopoulos, J. Brooks and T. Bunn for technical assistance. This work was supported by the NIDA and NIMH.

Correspondence and requests for materials should be addressed to R.E.H. (e-mail: rhampson@wfbumc.edu) or to S.A.D. (e-mail: sdeadwyl@wfbumc.edu).

Conversion of p35 to p25 deregulates Cdk5 activity and promotes neurodegeneration

Gentry N. Patrick^{*}, Lawrence Zukerberg^{*†}, Margareta Nikolic^{*‡}, Suzanne de la Monte^{†§}, Pieter Dikkes^{||} & Li-Huei Tsai^{*¶}

^{*} Department of Pathology, Harvard Medical School and [¶]Howard Hughes Medical Institute, 200 Longwood Avenue, Boston, Massachusetts 02115, USA

[†] Department of Pathology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

[§] Massachusetts General Hospital—East Cancer Center, Charlestown, Massachusetts 02119, USA

^{||} Department of Neurology, Children's Hospital, 300 Longwood Avenue, Boston, Massachusetts 02115, USA

[‡] Present address: Molecular Neurobiology Group, King's College London, London SE1 9RT, UK

Cyclin-dependent kinase 5 (Cdk5) is required for proper development of the mammalian central nervous system. To be activated, Cdk5 has to associate with its regulatory subunit, p35. We have found that p25, a truncated form of p35, accumulates in neurons in the brains of patients with Alzheimer's disease. This accumulation correlates with an increase in Cdk5 kinase activity. Unlike p35, p25 is not readily degraded, and binding of p25 to Cdk5 constitutively activates Cdk5, changes its cellular location and alters its substrate specificity. *In vivo* the p25/Cdk5 complex hyperphosphorylates tau, which reduces tau's ability to associate with microtubules. Moreover, expression of the p25/Cdk5 complex in cultured primary neurons induces cytoskeletal disruption, morphological degeneration and apoptosis. These findings indicate that cleavage of p35, followed by accumulation of p25, may be involved in the pathogenesis of cytoskeletal abnormalities and neuronal death in neurodegenerative diseases.

Cdk5 is a small protein serine/threonine kinase with close structural homology to the mitotic CDKs^{1,2}. Association of Cdk5 with p35, its regulatory subunit, is critical for kinase activation^{3–5}. The p35/Cdk5 kinase is required for neurite growth, as inactivation of Cdk5 in cultured cortical or cerebellar neurons inhibits neurite outgrowth^{6,7}. Mice lacking p35 display defects in cortical lamination and fasciculation of axon fibres^{8–10}. Interestingly, in p35^{−/−} cortex, the layering of cortical neurons is inverted, presumably because neurons born later cannot migrate past their predecessors⁹. Cdk5-null mice are not viable past postnatal day 0 (ref. 11). These animals have extensive abnormalities in the CNS including cortical and hippocampal lamination defects and cerebellar hypoplasia¹². These studies underscore the critical role of the p35/Cdk5 kinase in CNS development. Cdk5 activity is also required in the mature nervous system. Phosphorylation of DARPP32 by Cdk5 makes it an inhibitor of protein kinase A and alters the responses of striatal neurons to dopamine¹³. Cdk5 also phosphorylates Munc18, which in turn affects synaptic vesicle exocytosis¹⁴.

Alzheimer's disease (AD) is a degenerative disease characterized by progressive loss of neurons; its principal clinical manifestation is dementia. The main pathological features of AD include amyloid plaques, which are deposited extracellularly, and cytoplasmic fila-

mentous materials known as neurofibrillary tangles (NFT), which accumulate in the neuronal cell soma. Many studies have led to the conclusion that aberrantly phosphorylated forms of the microtubule-associated protein tau are the main constituents of NFT in the brains of patients with AD^{15,16}. Here we show that deregulation of Cdk5 contributes to the pathogenesis of neurodegenerative diseases such as AD. Deregulation of Cdk5 is caused by the accumulation of a truncated fragment of p35, p25, which is produced and accumulates in the brains of patients with AD. p25 causes Cdk5 to be constitutively activated and mislocalized *in vivo*. The p25/Cdk5 kinase phosphorylates tau efficiently and reduces the ability of tau to bind to microtubules. Moreover, p25/Cdk5 causes morphological degeneration and profound apoptotic cell death of primary neurons. These observations indicate that the conversion of p35 to p25 is involved in the pathogenesis of AD.

p25 accumulates in the brains of AD patients

We surveyed the expression profiles of p35 and Cdk5 from human brain tissues (Table 1). Whereas p35 levels were similar in all samples, a species with a relative molecular mass of 25,000 ($M_r=25K$), recognizable by anti-p35 antibodies, was found to be accumulated 20–40-fold (arrowhead) compared to p35 in all but one of the samples from patients with AD (Fig. 1a). Cdk5 levels do not vary significantly between normal and AD samples. Table 1 summarizes the cases used in this study. Figure 1a (lane 8) represents a sample from a patient in the terminal stages of AD with a substantial loss of brain tissue (brain weight of 950 g), which may account for the lack of accumulation of the 25K species. To verify the identity of the 25K species, we used antibodies recognizing various regions of p35 (Fig. 1c). p35 is the most prominent protein recognized by these antibodies in rat brain lysate, as detected by western blot analysis (Fig. 1d). The accumulation of this 25K species in samples from AD patients corresponded to the elevated Cdk5 kinase activity in the AD samples, as indicated by the Cdk5-associated histone H1 kinase activity (Fig. 1b). p35 is phosphorylated in a Cdk5 immunoprecipitation/kinase assay; the 25K species was found to be similarly phosphorylated (Fig. 1b). This 25K species was contained in immunocomplexes that were immunoprecipitated

Table 1 Case details

Age (years)	Sex	Diagnosis	Postmortem delay (h)	Lane
89	N/A	Control	<18	1
66	M	Control	<24	2
45	F	Control	<24	3
72	F	AD	3	4
59	F	AD	<18	5
71	F	AD	<18	6
64	F	AD	4.5	7
73	F	AD	14	8
66	N/A	Control	<18	9
N/A	N/A	HD	<24	10
78	F	AD	<18	11
68	M	AD	<18	12
87	M	AD	<18	13

N/A = not available

with either anti-Cdk5 antibodies (Fig. 1b) or antibodies recognizing the carboxy-terminal portion of p35 (Fig. 1e, right), indicating that the 25K species is either a cleavage product of p35 or derived from a related protein. Amino-terminal-specific anti-p35 antibodies did not recognize the 25K species (Fig. 1e, left). The size of the 25K species is reminiscent of p25, which was previously copurified with Cdk5 from bovine brain lysates^{3–5,17}. Indeed, the 25K species comigrated with p25 from COS-7 cell lysates transfected with cytomegalovirus (CMV)-p25 (a p35 N-terminal deletion mutant generated according to the published sequences) on SDS-polyacrylamide gel electrophoresis (SDS-PAGE) (Fig. 1e, right). To determine whether the 25K species observed in brain lysates from AD patients is actually p25, the V8 digest pattern of the 25K species from AD tissue was compared to that of p25 derived from COS-7 cell lysate transfected with CMV-p25. The two were identical (Fig. 1f). On the basis of the observations that the 25K species is recognized by the p35 C-terminal-specific antibodies but not by the

N-terminal-specific antibodies, that it associates with Cdk5 and with Cdk5 kinase activity, that it comigrates with p25 on SDS-PAGE and that it displays an identical V8 digestion pattern to that of recombinant p25, we conclude that the 25K species that accumulates in the brains of AD patients is indeed p25. As p35 is encoded by one exon, without interruption by introns, alternative splicing cannot account for the production of p25; rather, it is probably generated by proteolytic cleavage at a specific site (Fig. 1c).

p25 is found in neurons containing NFT in AD brain

As the levels of p25 protein were elevated in brain lysates from AD patients, it was of interest to determine the distribution of p25 immunoreactivity in histological sections from AD and control brain. The p25 polyclonal antibody labelled certain neurons in sections taken from the hippocampal formation of patients with AD (Fig. 2b). This antibody otherwise labelled the periphery of the perikaryon of hippocampal neurons in both AD and control sections (Fig. 2a). Adjacent sections were stained with AT8 (ref. 18), an anti-phospho-tau antibody, which labelled neurons containing NFT in the brains of AD patients (Fig. 2d) but not in control

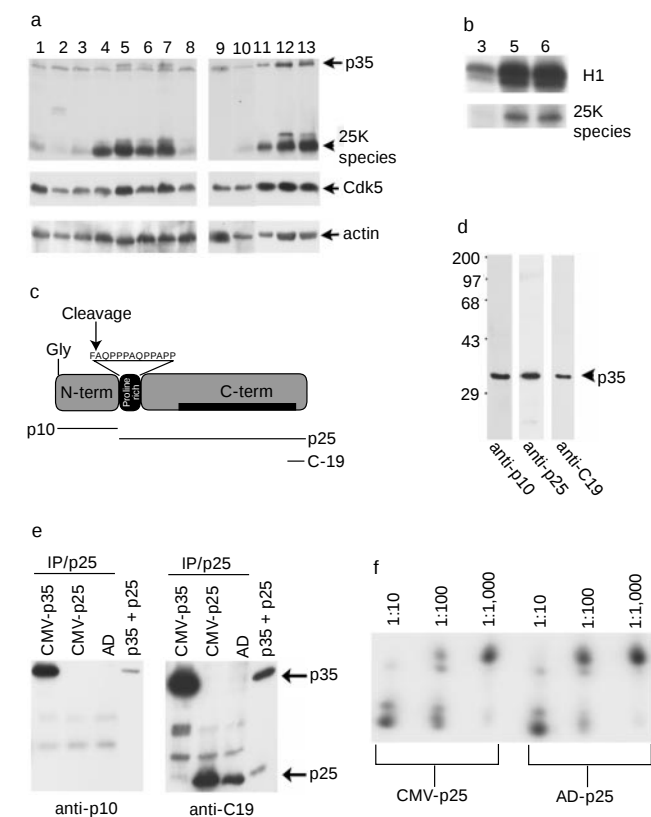


Figure 1 Accumulation of p25 in brain lysates from patients with AD. **a**, p35, Cdk5 and actin western blots of brain lysates from normal control (lanes 1–3 and 9) and AD tissues (lanes 4–8 and 11–13). Lane 10 is from a Huntington's disease patient. **b**, Cdk5 immunoprecipitation (IP)/kinase assays of control (lane 3) and AD (lanes 5 and 6) brain lysates using Cdk5 antibodies. Histone H1 was used as exogenous substrate. The 25K species was also immunoprecipitated and phosphorylated in the Cdk5 IP/kinase assay. **c**, Structure of p35. The N- and C-terminal regions of p35 are separated by a proline-rich region (black box); the minimal Cdk5 binding and activation domain is marked by thin black bar within the C-terminal region; cleavage between the phenylalanine and alanine residues liberates a 25K species. C-19 and p25 polyclonal antibodies recognize p35 and p25. p10 polyclonal antibodies recognize p35 only. **d**, Western blot analysis of P0 rat brain lysate probed with p10, C-19 and p25-specific antibodies. Left, $M_r (\times 10^3)$. **e**, p25 antibody immunoprecipitates of lysates from COS-7 cells transfected with CMV-p35 or CMV-p25 and AD brain, and western blot with either p10 N-terminal-specific or C-19 C-terminal-specific antibodies; p35 + p25 represents a mixture of CMV-p35- and CMV-p25-transfected lysates. **f**, V8 protease-digested fingerprints comparing recombinant p25 to the 25K species (previously immunoprecipitated and phosphorylated in Cdk5 IP/kinase reactions) identified in AD brain lysates.

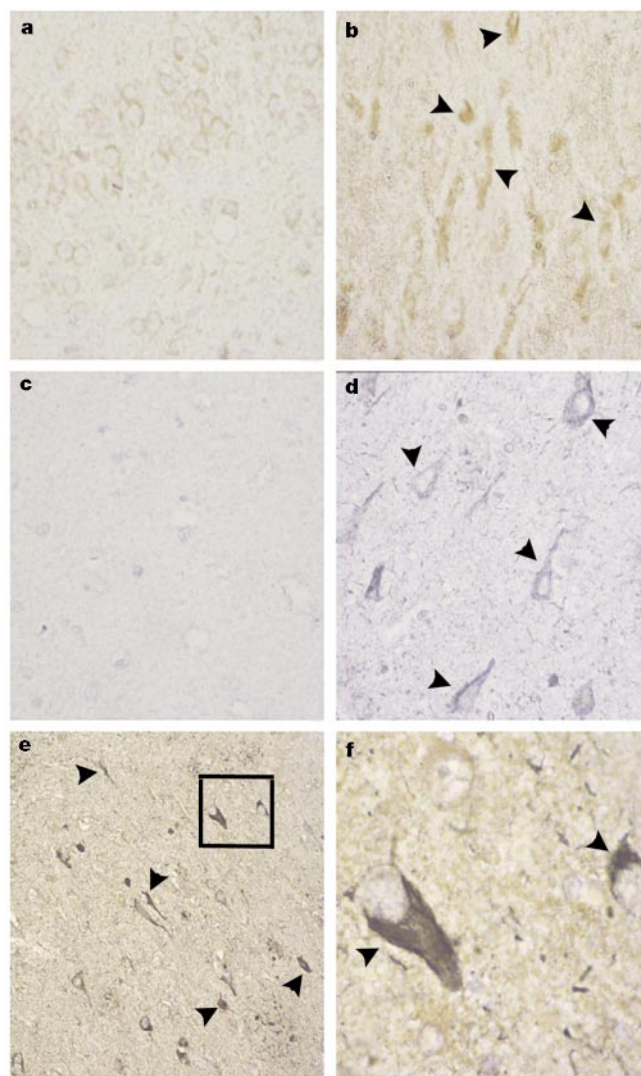


Figure 2 Immunohistochemistry of normal control and AD hippocampal sections. **a**, Control labelled with p25 antibody. **b**, AD labelled with p25 antibody (brown, arrowheads). **c**, Control labelled with AT8 (blue-grey). **d**, AD labelled with AT8 (blue-grey, arrowheads). **e**, AD labelled with p25 antibody (brown) and AT8 (blue-grey) detecting neurons containing NFT; double labelling appears blue-black (arrowheads). **f**, Inset from **e**; neurons double-labelled with p25 and AT8 antibodies (arrowheads). Original magnifications are 200x for **a–d**, 100x for **e** and 500x for **f**.

sections (Fig. 2c). Double immunostaining with anti-p25 and AT8 revealed that many NFT-containing neurons were also positive for p25 (Fig. 2e, f). However, there were more neurons with elevated p25 immunoreactivity than neurons displaying NFT (Fig. 2b, d). We also carried out immunostaining with anti-p10 antibodies, which detect only full-length p35 (Fig. 1c). Anti-p10 also labelled the neuronal periphery in control sections; its immunoreactivity did not accumulate in neurons containing NFT (data not shown). Together, western blot analysis and immunohistochemistry of AD tissues indicate that p25, but not p35, may accumulate in the brains of patients with AD and be present in neurons affected by neurodegeneration.

p25 differs from p35 in distribution and stability

p25 contains all the elements required for Cdk5 activation (Fig. 1c) and activates Cdk5 *in vitro*^{3,19}. As p25, but not p35, accumulates in AD, we attempted to determine whether these two proteins have different biochemical properties. Both Cdk5 and p35 are enriched in the processes and growth cones of neurons and p35 segregates with the plasma membrane^{6,20}. Figure 3a shows an N-terminal myristoy-

lation signal motif that is highly conserved in mammalian, *Xenopus*²¹ and *Caenorhabditis elegans* p35 homologues. To determine whether this myristoylation signal is required for the normal subcellular localization of p35, the conserved glycine at position 2 (Fig. 3a) was mutated to alanine and expressed in COS-7 cells (Fig. 3b). Although p35 was localized at the cell periphery and induced lamellipodia and filopodia (Fig. 3b, left; and ref. 20), the G2A mutant was absent at the cell periphery (Fig. 3b, right), showing that the myristoylation signal is essential for the proper distribution of p35. As p25 lacks the conserved myristoylation sequence we compared its subcellular localization with that of p35 in transfected fibroblasts. Figure 3c shows that p25 was enriched in nuclear and perinuclear regions of the cell, whereas p35 and p10 (an N-terminal fragment of p35 lacking the p25 region) were localized and enriched, in the case of p10, at the cell peripheries. We further investigated the distribution of p25 by subcellular fractionation. Figure 3d shows that, although p35 was more abundant in the membrane fraction, p25 was enriched in the cytosolic fraction. This may indicate that p35 is normally targeted to the membrane *in vivo*. In contrast, p25, which is not targeted to the plasma membrane, probably sequesters Cdk5 away from compartments of the cell where p35/Cdk5 activity is normally required. In addition, in primary cortical neurons, p25 is primarily concentrated in the cell soma and is largely absent from neurites (see below), whereas p35 is present in the peripheries and nerve terminals^{6,20}.

In addition to the difference in subcellular localization of p35 and p25, there is a marked difference in turnover rate between the two. Our previous studies indicated that Cdk5 activity is tightly regulated by p35 protein levels²². p35 has a half-life ($t_{1/2}$) of ~20–30 min in primary cortical neurons; this rapid turnover rate is in part due to phosphorylation-stimulated, ubiquitin-mediated degradation²². When expressed in COS-7 cells together with Cdk5, p25 had a ~5–10-fold longer half-life than p35 or the p35 G2A mutant (Fig. 4a, b). Cdk5-associated histone H1 kinase activity paralleled the levels of p35 or p25. These observations indicate that the accumulated p25 in the brains of AD patients may cause prolonged activation of Cdk5 and mislocalization of Cdk5 kinase activity in affected neurons.

Tau phosphorylation by the p25/Cdk5 kinase *in vivo*

To explore further the functional difference between the p35/Cdk5 and p25/Cdk5 kinases *in vivo*, we compared the activity of these two

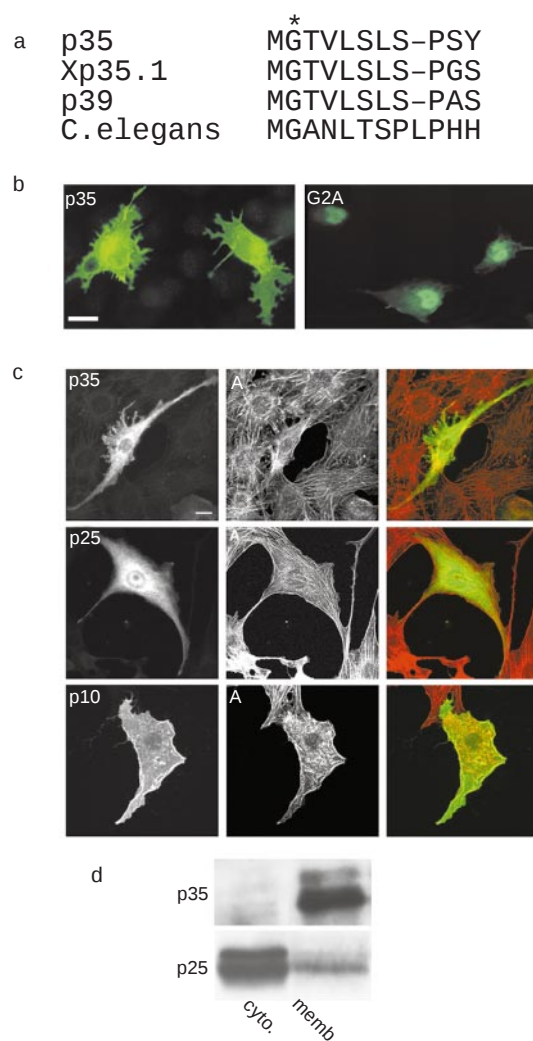


Figure 3 p35 but not p25 is associated with peripheral membranes. **a**, Myristoylation sequence within p35 homologues is conserved. **b**, COS-7 cells were transfected with either wild-type p35 or the p35 G2A mutant followed by p35 immunostaining. **c**, Confocal immunofluorescence of Swiss 3T3 cells transfected with p35, p25 or p10-HA (overlay: red, rhodamine-conjugated phalloidin, labelled A for actin; green, p35, p25 or p10-HA). All scale bars are 10 μ m. **d**, Subcellular distribution of p35 and p25 in fractionated COS-7 cell lysates after transfection.

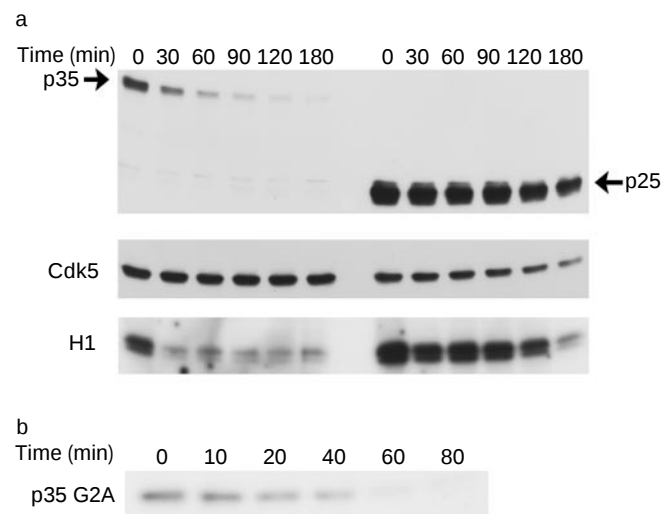


Figure 4 p25 is a stable protein. Western blots of lysates from cycloheximide treatment half-life experiments of p35 and p25 (**a**) and the p35 G2A mutant (**b**) in transfected COS-7 cells. Also in **a**, western blot analysis of Cdk5 and *in vitro* Cdk5-associated histone H1 kinase assays from lysates used in cycloheximide treatment half-life experiments.

kinases in phosphorylation of the microtubule-associated protein tau. Tau is a substrate of Cdk5 (ref. 23). A His₆-tagged human tau 40 (his₆-htau40) was cotransfected with p35/Cdk5, p25/Cdk5 or p25/DNK5 (a catalytically inactive mutant of Cdk5) in COS-7 cells. Tau phosphorylation was evaluated by immunoblotting with AT8 or PHF-1. AT8 recognizes tau epitopes phosphorylated at Ser 202 and Ser 205 and PHF-1 recognizes epitopes phosphorylated at and around Thr 396, sites previously shown to be phosphorylated by Cdk5 *in vitro*^{24–27}. Intense AT8 immunoreactivity was seen in cells expressing p25/Cdk5 (Fig. 5a, lane 5) but not p35/Cdk5 or p25/DNK5 (Fig. 5a, lanes 3 and 6). Interestingly, PHF-1 detected a slower migrating species in the p25/Cdk5 expressing cells (Fig. 5a, lane 5 and 5b, arrowhead) which was readily abolished upon protein phosphatase treatment (Fig. 5a, lanes 8 and 9), indicating that this slower migrating band was indeed a phosphorylated species of tau. This species was not present in cells expressing p35/Cdk5 or p25/DNK5. As a control, Fig. 5a (lane 7) shows that GSK3- β , a well established kinase for tau^{28,29}, caused a large increase in tau PHF-1 immunoreactivity.

Owing to the difference in turnover rate, the steady-state levels of p35 were always much lower than those of p25 (Fig. 5a), which may have contributed to the observed difference in tau phosphorylation. To compare the ability of these kinases to phosphorylate tau when similar levels of the two kinases are expressed *in vivo*, we used five times more p35 than p25 plasmid DNA for transfection. However, even after p35 and p25 were expressed at comparable levels, the slower migrating PHF-1 immunoreactive species of tau was still absent from p35/Cdk5 transfected cells. This is despite the observed increase in PHF-1 immunoreactivity (Fig. 5b), indicating that other differences such as the altered subcellular distribution of the p25/Cdk5 kinase may allow more efficient targeting of tau *in vivo*. Comparable levels of tau were expressed, as indicated by immunoblotting with anti-His₆ antibodies (Fig. 5a). Anti-His₆ antibodies also revealed a shift in tau mobility in cells cotransfected with p25 and Cdk5 (Fig. 5a, lane 5). Transfection of p25 alone did not cause a noticeable increase in PHF-1 or AT8 immunoreactivity, which is likely to reflect the low endogenous level of Cdk5 in COS-7 cells (Fig. 5a, lane 4). When transfected COS-7 cell lysates were incubated with polymerized microtubules, tau's ability to bind microtubules was impaired in cells co-expressing the active p25/Cdk5 kinase complex (Fig. 5c), indicating that the hyperphosphorylation of tau by p25/Cdk5 may affect its function *in vivo*.

p25/Cdk5 causes cytoskeletal disruption

The effects of p25/Cdk5 on tau hyperphosphorylation were also investigated in primary cortical neurons using a herpes simplex recombinant viral expression system. After infection with a β gal virus, weak PHF-1 immunoreactivity was present in the cell soma and axon fibres (Fig. 6a, d). In contrast, infection with p25/Cdk5 produced robust PHF-1 immunoreactivity which was concentrated in the cell soma (Fig. 6b, e, h, i). This effect was reversed when p25/DNK5 was expressed (Fig. 6c, f), indicating that the catalytic activity of Cdk5 is necessary for the observed increase in PHF-1 signal. Neurons infected with p35/Cdk5 also displayed increased PHF-1 immunoreactivity, but to a lesser extent (Fig. 6g). Similar results were obtained using AT8 antibodies (Fig. 6j–k). Many p25/Cdk5-infected neurons, indicated by PHF-1-positive staining, exhibited neurite retraction (Figs 6h, i, 7b–c and f–g), whereas cultures infected with p35/Cdk5 seldom exhibited signs of neurite degeneration (Fig. 6g). Neurofilaments are also well established substrates of Cdk5 (refs 1, 3, 30–32). Using SMI34, a phospho-specific neurofilament H antibody, we observed intense phospho-neurofilament immunoreactivity in p25/Cdk5-infected neurons (Fig. 6m, n, arrowheads), but less intense staining in uninfected neurons (arrows), indicating that the p25/Cdk5 kinase may cause hyperphosphorylation of neurofilaments.

Tau hyperphosphorylation causes microtubule destabili-

zation^{16,33}. We have shown that p25/Cdk5-phosphorylated tau binds to microtubules less well. We therefore used a silver-based stain to assess the cytoskeletal integrity of p25/Cdk5-expressing neurons. Figure 6o shows that silver-positive neurons were frequently seen in the p25/Cdk5-expressing cultures (arrows). Silver labelling was specific to p25/Cdk5-infected neurons and was never observed in cultures infected with β gal, p35/Cdk5 or p25/DNK5 (data not shown). These data indicate that p25/Cdk5 may be more potent than p35/Cdk5 in phosphorylating protein substrates such as tau and neurofilaments *in vivo*, which may be, in part, attributable to its accumulation and different localization. Furthermore, they indicate that the presence and accumulation of the p25/Cdk5 kinase is associated with morphological degeneration and cytoskeletal disruption in neurons, events seen in AD and other neurodegenerative diseases.

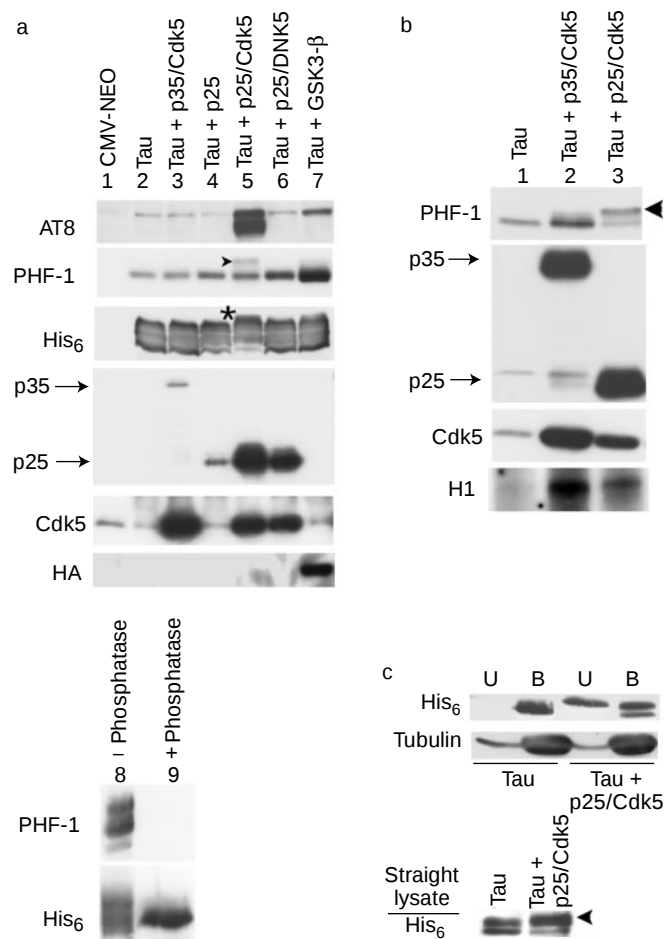


Figure 5 Comparison of tau phosphorylation by p35/Cdk5 and p25/Cdk5. **a**, COS-7 cell lysates transfected with empty vector (lane 1), his₆-htau40 alone (lane 2) or with p35/Cdk5 (lane 3), p25 (lane 4), p25/Cdk5 (lane 5, 8, and 9), p25/DNK5 (lane 6) or HA-tagged GSK3- β (lane 7) CMV expression constructs and probed with phospho-specific tau antibodies AT8 and PHF-1, and anti-His₆, p35, Cdk5 and 12CA5 (HA) antibodies. Arrowhead and asterisk indicate slower migrating phospho-species of tau. Lanes 8 and 9, immunoprecipitation of lysate from lane 5 with anti-His₆ and probed with PHF-1 and anti-His₆ after treatment with buffer alone (lane 8) or potato acid phosphatase (lane 9). **b**, As in **a**, except comparable levels of p35 and p25 were expressed in COS-7 cells after using 5-fold more p35 plasmid DNA; arrowhead indicates shift in tau phosphorylation by p25/Cdk5, absent in p35/Cdk5-transfected cells. **c**, Microtubule-binding assay of COS-7 cells transfected with tau and tau plus p25/Cdk5. U, tau unbound to microtubules; B, tau bound to microtubules. Whereas tau from cells transfected with tau alone all bound to microtubules, a large portion of tau from cells transfected with tau plus p25/Cdk5 was present in the unbound fraction (arrowhead denotes slower migrating species of tau).

The p25/Cdk5 kinase induces apoptosis in neurons

To characterize the cytoskeletal abnormalities further we stained for β -tubulin in the infected cultures. β -tubulin staining revealed a marked difference in the microtubule networks of p25/Cdk5-infected cells compared with those of p25/DNK5- or control β gal-infected cells (Fig. 7a–h). Tubulin was normally distributed throughout the cell soma and the axonal and dendritic compartments in β gal- and p25/DNK5-infected cultures (Fig. 7a, e and d, h, respectively). In contrast, tubulin was concentrated in the perikarya of p25/Cdk5-infected cells (Fig. 7f, g, arrowheads). In fact, most of the p25/Cdk5-infected cells were devoid of neurites. Apoptotic cell bodies were frequently observed in p25/Cdk5-infected cultures (Fig. 7f, arrows), but not in cultures infected with β gal or p25/DNK5 (Fig. 7e, h).

We also found in COS-7 cell transfection experiments that many cells died upon co-expression of p25 and Cdk5. We compared the extent of cell death in COS-7 cells transfected with p25/Cdk5 or

p35/Cdk5 by purifying soluble fragmented DNA from these cells. DNA laddering, which was evident in p25/Cdk5-transfected cells, was present at a much reduced level in cells transfected with p35/Cdk5 or with empty vector (control cells) (Fig. 7i). We examined the nuclear morphology of infected cells by Hoechst stain. β gal-infected neurons had normal nuclear morphology (Fig. 7j, m), whereas most p25/Cdk5-infected neurons had fragmented and condensed nuclei (Figs 6l, 7k, n). Moreover, most of the p25/DNK5-infected neurons had normal nuclear morphology (Fig. 7l, o). Neurons with fragmented nuclei were also positive for the TdT-mediated dUTP nick end labelling (TUNEL) assay (data not shown). In general, more than 85% of p25/Cdk5-infected neurons had fragmented or condensed nuclear morphology, and the effect was largely reversed by the expression of DNK5 (Fig. 8a). Very few neurons had nuclear fragmentation after β gal viral infection. About 30% of p35/Cdk5-infected neurons had disrupted nuclei. These results were corroborated by calcium phosphate transfection of primary cortical neurons. A β gal DNA construct was cotransfected with various plasmid DNAs at a ratio of 1:5 to ensure that most positively scored β gal cells expressed the genes of interest. About 65–70% of p25/Cdk5-transfected neurons had fragmented nuclei, whereas less than 5% of neurons transfected with β gal alone and around 20% of those transfected with p35/Cdk5 had abnormal nuclear morphology (Fig. 8b). In addition, the p35 G2A mutant did not cause significant cell death (Fig. 8b). GSK3- β expression caused no detectable abnormalities in nuclear morphology in primary cortical neurons. The p25/Cdk5-induced nuclear condensation/fragmentation could be partially inhibited by Ac-DEVD-CHO, an inhibitor of caspase-3 (Fig. 8c).

The apoptotic cell bodies revealed by tubulin staining, DNA

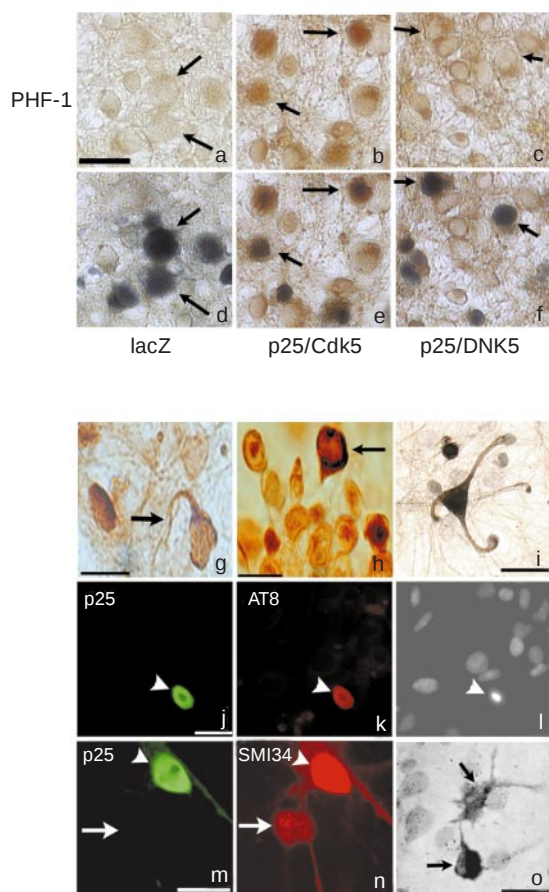


Figure 6 Effects of the p25/Cdk5 kinase on the neuronal cytoskeleton. Primary cortical cultures infected with β gal (a, d), p35 and Cdk5 (g), p25 and Cdk5 (b, e, h, i, m–o) and p25 and DNK5 (c, f) recombinant HSV viruses. Immunostaining for phospho-tau with PHF-1 (brown, a–c and g–i). d–f, Double labelling for either β gal (d) or p25 (e, f) to reveal infection-positive cells (dark blue). Arrows indicate positively infected neurons. g–i, High magnification of neurons infected with p35/Cdk5 (g, arrow labels neurite) and p25/Cdk5 (h, arrow labels neuron with enriched tau immunoreactive material in cell soma). i, p25/Cdk5-infected neuron with degenerated neurites, double labelled with PHF-1 (brown) and p25 antibodies (blue). j–l, Increased AT8 immunoreactivity in p25/Cdk5-transfected neurons. j, β gal-positive neuron (arrowhead) indicating transfected neuron. k, AT8 labelling of positively transfected neuron (arrowhead). l, Hoechst dye labelled nuclei depicting condensed nuclei (arrowhead) in p25/Cdk5-transfected neurons as compared to nontransfected cells. m, n, Phospho-neurofilament H (SMI34) fluorescent staining in n and p25 staining in m of p25/Cdk5-infected (arrowheads) and uninfected (arrows) neurons. o, Silver stain of p25/Cdk5-infected cultures (arrows). All scale bars are 10 μ m.

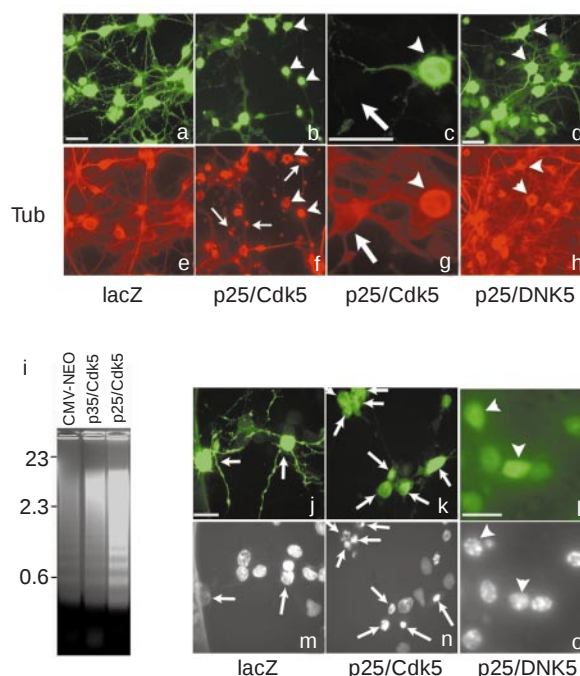


Figure 7 The p25/Cdk5 kinase induces cytoskeletal collapse and nuclear fragmentation. a–h, Cytoskeletal collapse; j–o, nuclear fragmentation. Fluorescent staining of neurons infected with β gal, p25/Cdk5 and p25/DNK5; β gal (green, a and j) and p25 (green, b–d and k, l), β -tubulin (red, e–h), or with Hoechst dye (m–o). In b, c, 2f and g arrowheads indicate neurons devoid of most neurites; arrowheads in d and h indicate p25/DNK5-infected neurons; arrows in f indicate tubulin fragments (similar to apoptotic bodies); arrows in c and g indicate non-infected neurons. Arrows in k and n indicate infected neurons with apoptotic nuclear morphology; arrows in j and m, and arrowheads in l and o, indicate β gal- or p25/DNK5-infected neurons with normal nuclei. Scale bars are 10 μ m. i, Soluble fragmented DNA purified from mock-, p35/Cdk5- and p25/Cdk5-transfected COS-7 cells.

laddering and fragmented and condensed nuclear morphology strongly support a pro-apoptotic effect for the p25/Cdk5 kinase. The inability of the catalytically inactive Cdk5 mutant to induce apoptosis indicates that substrate phosphorylation may be necessary for cell death. The reduction in apoptotic cell death resulting from p35/Cdk5 overexpression supports the idea that deregulation of Cdk5 is detrimental to cells. It further indicates that, in addition to the increase in kinase level, p25-mediated mislocalization of

Cdk5 activity also contributes to the observed degeneration in neurons.

Discussion

We have shown that a proteolytic fragment of p35, p25, accumulates in the brains of AD patients and is present in neurons with NFT. Whereas p35 is required for normal brain development, the presence of p25 causes deregulation of Cdk5 kinase activity, due in part to the fact that p25 is a stable protein and that it is inappropriately localized. Other evidence indicates that the N-terminal portion of p35 may be necessary for binding to regulatory proteins³⁴. Therefore, it is conceivable that, in neurons containing high levels of p25, Cdk5 is sequestered from normal regulation and concentrated at an abnormal site(s), and phosphorylates substrates not normally phosphorylated (or hyperphosphorylated) by this kinase. For instance, we show here that the p25/Cdk5 kinase displays an increased and altered tau phosphorylation in comparison to the p35/Cdk5 kinase *in vivo*. Neurofilament H is also heavily phosphorylated in p25/Cdk5-expressing neurons. Many reports indicate that hyperphosphorylation of microtubule-associated proteins such as tau alters their interactions with microtubules and causes microtubule instability. Indeed, tau phosphorylated by the p25/Cdk5 kinase displays reduced binding to microtubules. In addition, p25/Cdk5-expressing neurons showed neurite retraction and signs of microtubule collapse. Many of them could be labelled with a silver-based staining method which is commonly used as an indicator of cytoskeletal disruption. Thus, deregulation of Cdk5 by the accumulation of p25 impairs the integrity of the cytoskeleton, which ultimately results in morphological degeneration and, perhaps, neuronal apoptosis. Morphological degeneration and neuronal death are fundamental aspects of many neurodegenerative diseases. Our findings are nicely corroborated by the finding that the p35/Cdk5 kinase is associated with neuronal death^{35–38}. Other possible deleterious effects of p25 are that it may prevent the p35/Cdk5 kinase from phosphorylating its normal substrates, such as Pak1, synapsin, syntaxin, Munc18 and DARP32 (refs 13, 20, 39 and 40) by sequestering Cdk5 from the cell periphery and nerve terminals, which also contributes to neuronal dysfunction.

We propose that conversion of p35 to p25 results in deregulation of the Cdk5 kinase. The deregulated Cdk5 kinase can cause irreversible damage to the cytoskeleton and neuronal death (Fig. 8d). On the basis of the accumulation of p25 in the brains of patients with AD and the cytoskeletal disruption and apoptosis induced by the p25/Cdk5 kinase in neurons, we suggest that p25 production and accumulation in the brain may contribute to the pathogenesis of AD. p25 may contribute to the early stages of AD, as it was not found to be accumulated in a terminal stage AD patient with significant cell loss. Furthermore, neurons with elevated p25 outnumbered neurons containing NFT in sections from the brains of AD patients. Thus, p25 accumulation may precede the formation of NFT. Conversion of p35 to p25 is likely to be the consequence of proteolytic cleavage; our preliminary results indicate that cleavage of p35 to p25 can be activated upon oxidative stress (Y. T. Kwon, M. S. Lee and L.-H.T., unpublished data). It will be of interest to identify the putative protease that cleaves p35, as it may serve as a therapeutic target for prevention and treatment of neurodegenerative diseases.

Methods

Chemicals and antibodies

Cycloheximide (20 mg ml⁻¹ stock; Sigma) was used at a final concentration of 30 µg ml⁻¹ in *t*_{1/2} experiments. Hoechst dye was also purchased from Sigma. The following antibodies were used: p35: polyclonal antibodies (pAb) neu-cyc (purified with either glutathione *S*-transferase (GST)–p10 (N-terminal—GST–p10 purified) or GST–p25 (C-terminal—GST–p25 purified)), 4E3 raised against whole protein⁶, N-20 and C19 (Santa Cruz); Cdk5: monoclonal antibodies (mAb) DC17 (ref. 41), pAb C8 (Santa Cruz); haemagglutinin (HA): mAb 12CA5; mAb his₆ (Boehringer Mannheim); pAb βgal antibodies (Promega); pAb actin; mAb α- and β-tubulin (Sigma); tau: dephosphorylated tau epitopes at Ser 199

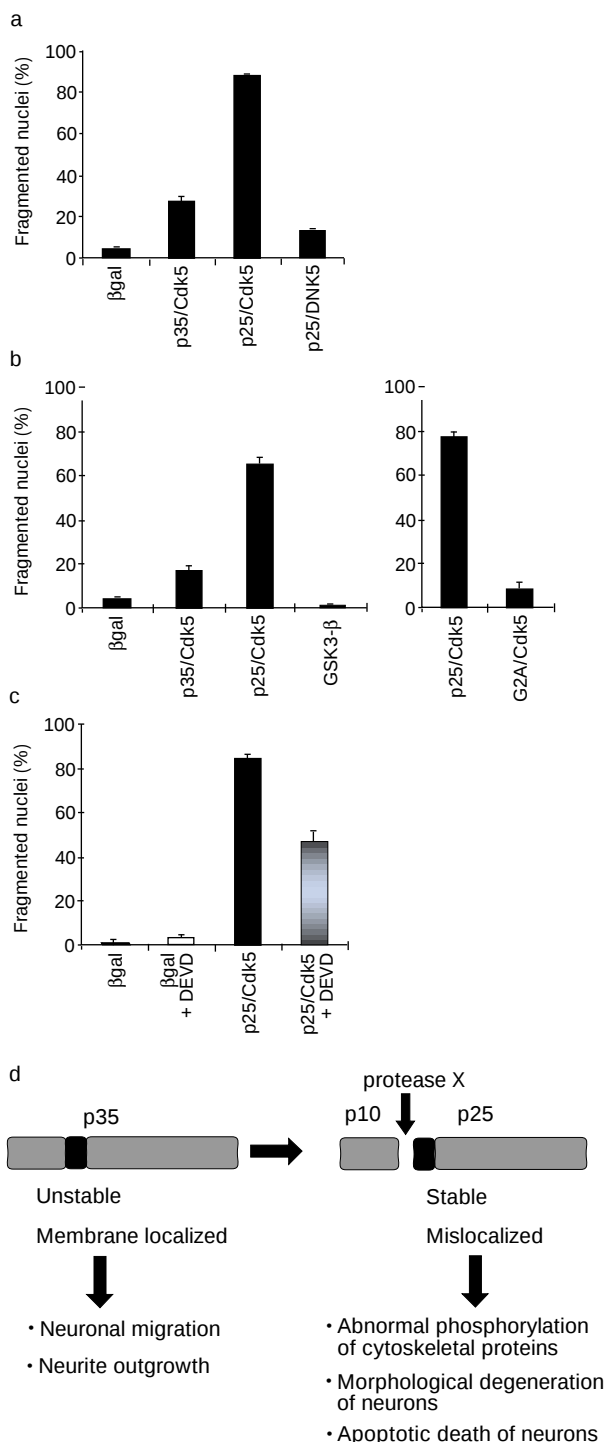


Figure 8 Quantification of nuclear fragmentation induced by the p25/Cdk5 kinase. **a–c**, Per cent fragmented nuclei of HSV-infected (**a**), calcium phosphate-transfected (**b**) and HSV-infected plus treated with DEVD inhibitor (**c**) cortical cultures. **d**, Model: cleavage of p35 and accumulation of p25 in neurons causes neuronal degeneration.

and Ser 202—mAb Tau-1 (Boehringer Mannheim)⁴²; phosphorylated epitopes at Ser 396 and Ser 404—mAb PHF-1 (a gift from P. Davies)⁴³; phosphorylated epitopes at N-terminal residues 2–10—mAb Alz-50 (a gift from P. Davies)⁴⁴; AT8 (Innogenetics). The caspase-3 inhibitor Ac-DEVD-CHO was purchased from Sigma.

Constructs and viruses

Construction of CMV expression vectors for p10, p35, Cdk5 and DNK5 was as described⁵. CMV–p25 (amino acids 98–307) was made by polymerase chain reaction (PCR) using a 3' p35 primer and 5' deletion primer. CMV–p10 (amino acids 1–97) was made by PCR using a 5' p35 primer and a 3' deletion primer. The p10 fragment was then cloned upstream of a haemagglutinin (HA) tag. p35 Gly2Ala was made by site-directed mutagenesis²⁹ using the following oligo: 5'-CAGACACCATGGCCACGGTG-3'. Mutagenesis was verified by sequencing. Recombinant herpes simplex virus constructs (a gift from R. Neve) were made as described⁴⁵. The following viruses were used: HSV– β gal, HSV–p35, HSV–p25, HSV–GFP–Cdk5 and HSV–GFP–DNK5 (where GFP is green fluorescent protein). Multiplicity of infection ranged from 0.1 to 1 for all viruses used. The HA–GSK3- β mammalian expression construct was a gift from X. He; the his₆–htau40 mammalian expression construct was a gift from D. Auprin (Pfizer).

Tissues and cell culture

We used brain tissue from eight patients with AD, four age-matched non-neurological cases and one patient with Huntington's disease. Brain lysates were made from Brodmann areas 11, 21 or 45 by Dounce homogenizing tissue in lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1% Triton-X, 10% glycerol, 5 mM EDTA, 1 mM EGTA, 1 mM DTT) plus inhibitors (2 μ g ml⁻¹ aprotinin, 2 μ g ml⁻¹ leupeptin, 1 μ g ml⁻¹ pepstatin, 50 mM β -glycerophosphate, 5 mM NaF, 5 mM NaVO₃ and 100 μ g ml⁻¹ PMSF). Lysates were cleared by centrifugation at 13,000g for 30 min. E17–19 pregnant rats (Long Evans strain) were purchased from Harland Sprague-Dawley. P0 pups were killed and their cortices were dissected and cultured as described⁶. Cortical cultures were grown in 24-well plates on either plastic or glass coverslips (400,000 cells per well) that had been treated with laminin and poly-D-lysine. Cultures were maintained in basal growth medium (BGM) for three days before viral infection. Swiss 3T3 and COS-7 cells were maintained in DMEM supplemented with 10% fetal calf serum. Ac-DEVD-CHO was added to medium (10 μ M) ~8 h before infection and then every day (10 μ M) after infection until cells were collected.

Transfection of COS-7 cells, cycloheximide treatment and kinase assays

COS-7 cells were transiently transfected with various plasmid constructs using calcium phosphate transfection procedures. For Fig. 3a, amounts of CMV plasmid DNA used were as follows: his₆–htau40 (5 μ g), p35 (5 μ g), p25 (2.5 μ g), Cdk5 (5 μ g), DNK5 (5 μ g) and HA–GSK3- β (10 μ g); for Fig. 3b we used five times as much p35 as p25 plasmid DNA. Cycloheximide *t*_{1/2} experiments were performed as described²². Histone H1 kinase activity was determined as follows: ~1 mg of protein from brain lysates or transfected cell lysates was immunoprecipitated with p35 (4E3) or Cdk5 (C8) antibodies, respectively. Histone H1 was added as a substrate in the *in vitro* kinase assay performed as described⁴¹. In the case of p25 phosphorylation by Cdk5, no histone H1 was added. Subcellular fractionation was done as described^{6,20}.

Immunohistochemistry

P0 cortical cultures were infected with various combinations of HSV– β gal, HSV–p35, HSV–p25, HSV–GFP–Cdk5 and HSV–GFP–DNK5 after three days in culture. Two to three days after infection, cultures were fixed with acetone:methanol (1:1) for 3 min at room temperature (RT), washed three times with PBS and permeabilized with 0.2% Triton X-100 in blocking solution (1% blocking reagent (Boehringer Mannheim) in PBS) for 30 min at RT. Cultures were washed three times with PBS and incubated with either PHF-1 (1:100), AT8 (1:100), p35 (1:400) or β gal (1:400) antibodies for 12 h in blocking solution. The cultures were then washed three times with PBS, and rabbit or mouse secondary biotinylated antibodies were added (1:150) in blocking solution and detected by the Vector Elite substrate kit (Vector). Diaminobenzidine (DAB—brown) and Vector SG substrate (blue) were used in double-labelling experiments. Double labelling was performed sequentially.

Calcium phosphate transfection of cortical cultures (E17–19; after two days in culture) was performed as described⁶. A CMV– β gal plasmid (10 μ g) was transfected with a CMV–NEO control plasmid (50 μ g), CMV–p35 and CMV–Cdk5 (25 μ g each), CMV–p25 and CMV–Cdk5 (25 μ g each) or CMV–HA–GSK3- β (50 μ g). Two to three days after transfection the cells were fixed and stained with β gal antibodies and Hoechst dye.

Transfected Swiss 3T3 cells were fixed with 4% paraformaldehyde and permeabilized in 0.3% Triton X-100. Immunostaining was done as described⁶ using anti-HA (mAb 12CA5), anti-p25 (pAb neu-cyc, pAb N-20; Santa Cruz) and anti-actin (mAb; Sigma), to detect p35, p25, p35 G2A, HA-p10 and actin. The cells were then treated with biotin-conjugated anti-rabbit (followed by FITC-conjugated streptavidin) and Texas-Red-conjugated anti-mouse antibodies. Coverslips were mounted in ProLong antifade (Molecular Probes) and analysed using a Leica or Zeiss confocal microscope. Where stated, normal fluorescence microscopy was used.

For staining of control and AD brain tissue, paraffin sections (8 μ m thick) were de-waxed in xylene, hydrated through graded alcohol solutions and blocked with 3% BSA, 10% NGS and 0.1% Triton X-100. The sections were incubated in citrate buffer for 10 min at 95 °C for antigen retrieval. The sections were incubated for 1 h at RT or overnight at 4 °C with 1–4 μ g ml⁻¹ of primary antibody. Bound rabbit and mouse antibodies were detected using Vectastain Elite Avidin-Biotin kit (Vector, Burlingame, CA) with DAB or Vector SG as substrate. For double-labelling studies, binding of the first primary antibody was

detected as described, and binding of the second primary antibody was detected using DAB or Vector SG as substrate. At the conclusion of the immunostaining reactions, the sections were dehydrated and mounted with Permount (Fisher Scientific) under cover-glass.

A modified silver stain was used as described⁴⁶. Infected cultures were fixed with 90% ethanol, 5% formaldehyde and 5% acetic acid before silver staining.

Microtubule-binding assay

The binding of tau to Taxol-stabilized microtubules (microtubules were assembled from purified bovine tubulin (Cytoskeleton) and stabilized by Taxol) was assayed as described⁴⁷. Briefly, ~1 mg of 200,000g cell lysates (lysed in 100 mM PIPES, pH 6.8, 0.1% Triton X-100, 1 mM MgCl₂, 1 mM EGTA, 1 mM GTP plus inhibitors) from transfected cells were incubated with 100 μ g Taxol-polymerized microtubules at 35 °C for 30 min. The microtubule-bound proteins were separated by centrifugation at 70,000g for 30 min. Pellets were resuspended in the same volume as supernatants and both bound and unbound fractions were analysed by western blot analysis.

Western blot analysis

P0 cortical cultures, brain lysates and transfected cells were washed with PBS and lysed with RIPA buffer plus inhibitors (2 μ g ml⁻¹ aprotinin, 2 μ g ml⁻¹ leupeptin, 1 μ g ml⁻¹ pepstatin, 5 mM NaF, 5 mM NaVO₃ and 100 μ g ml⁻¹ PMSF). In some cases, samples were immunoprecipitated with either p35 (polyclonal) or Cdk5 (C8) antibodies. Sample buffer was added to lysates and immunoprecipitates and samples were run on SDS–PAGE gels (see text and Figure 1 legends for gel percentage), electrotransferred to nylon membrane and probed with either Tau-1 (1:2,000), PHF-1 (1:1,000), AT-8 (1:1,000), β -tubulin (1:2,000), anti-His₆ (1:2,000), anti-HA (12CA5; 1:25), DC17 (1:20) or p35 (C19; Santa Cruz) (1:2,000), p35 GST–p25 purified polyclonal (1:2,000) and GST–p10 purified polyclonal (1:1,000).

Cell death measurements and DNA laddering

After immunostaining, neurons or COS-7 cells were labelled with DNA dye Hoechst 33258 (2.5 μ g ml⁻¹, 5 min), and infected neurons or transfected COS-7 cells were scored for healthy or apoptotic nuclear morphology. Cells were scored positive if they had a pyknotic and/or fragmented nucleus. Representative graphs are shown for experiments where 150 or 300 cells were scored. All experiments were done at least three times. TUNEL assays were done according to standard procedure (Boehringer Mannheim). For DNA laddering cells were lysed and soluble fragmented DNA was purified as described^{48,49}. DNA was run on a 1.5% agarose gel.

Received 26 August; accepted 12 October 1999.

- Lew, J., Winkfein, R. J., Paudel, H. K. & Wang, J. H. Brain proline-directed protein kinase is a neurofilament kinase which displays high sequence homology to p34cdc2. *J. Biol. Chem.* **267**, 25922–25926 (1992).
- Meyerson, M. *et al.* A family of human cdc2-related protein kinases. *EMBO J.* **11**, 2909–2917 (1992).
- Lew, J. *et al.* A brain-specific activator of cyclin-dependent kinase 5. *Nature* **371**, 423–426 (1994).
- Ishiguro, K. *et al.* Identification of the 23 kDa subunit of tau protein kinase II as a putative activator of Cdk5 in bovine brain. *FEBS Lett.* **342**, 203–208 (1994).
- Tsai, L. H., Delalle, I., Caviness, V. S. Jr, Chae, T. & Harlow, E. p35 is a neural-specific regulatory subunit of cyclin-dependent kinase 5. *Nature* **371**, 419–423 (1994).
- Nikolic, M., Dudek, H., Kwon, Y. T., Ramos, Y. F. & Tsai, L. H. The Cdk5/p35 kinase is essential for neurite outgrowth during neuronal differentiation. *Genes Dev.* **10**, 816–825 (1996).
- Paglini, G. *et al.* Evidence for the participation of the neuron-specific CDK5 activator P35 during laminin-enhanced axonal growth. *J. Neurosci.* **18**, 9858–9869 (1998).
- Kwon, Y. T., Tsai, L. H. & Crandall, J. E. Callosal axon guidance defects in p35–/– mice. *J. Comp. Neurol.* **415**, 218–229 (1999).
- Kwon, Y. T. & Tsai, L. H. A novel disruption of cortical development in p35(–/–) mice distinct from reeler. *J. Comp. Neurol.* **395**, 510–522 (1998).
- Chae, T. *et al.* Mice lacking p35, a neuronal specific activator of Cdk5, display cortical lamination defects, seizures, and adult lethality. *Neuron* **18**, 29–42 (1997).
- Ohshima, T. *et al.* Targeted disruption of the cyclin-dependent kinase 5 gene results in abnormal corticogenesis, neuronal pathology and perinatal death. *Proc. Natl Acad. Sci. USA* **93**, 11173–11178 (1996).
- Gilmore, E. C., Ohshima, T., Goffinet, A. M., Kulkarni, A. B. & Herrup, K. Cyclin-dependent kinase 5-deficient mice demonstrate novel developmental arrest in cerebral cortex. *J. Neurosci.* **18**, 6370–6377 (1998).
- Bibb, J. A. *et al.* Phosphorylation of DARPP-32 by Cdk5 modulates dopamine signalling in neurons. *Nature* **402**, 669–671 (1999).
- Fletcher, A. L. *et al.* Regulation of exocytosis by cyclin-dependent kinase 5 via phosphorylation of Munc18. *J. Biol. Chem.* **274**, 4027–4035 (1999).
- Mandelkow, E. M. & Mandelkow, E. Tau in Alzheimer's disease. *Trends Cell Biol.* **8**, 425–427 (1998).
- Spillantini, M. G. & Goedert, M. Tau protein pathology in neurodegenerative diseases. *Trends Neurosci.* **21**, 428–433 (1998).
- Uchida, T. *et al.* Precursor of Cdk5 activator, the 23 kDa subunit of tau protein kinase II: its sequence and developmental change in brain. *FEBS Lett.* **355**, 35–40 (1994).
- Biernat, J. *et al.* The switch of tau protein to an Alzheimer-like state includes the phosphorylation of two serine-proline motifs upstream of the microtubule binding region. *EMBO J.* **11**, 1593–1597 (1992).
- Poon, R. Y., Lew, J. & Hunter, T. Identification of functional domains in the neuronal Cdk5 activator protein. *J. Biol. Chem.* **272**, 5703–5708 (1997).
- Nikolic, M., Chou, M. M., Lu, W., Mayer, B. J. & Tsai, L. H. The p35/Cdk5 kinase is a neuron-specific Rac effector that inhibits Pak1 activity. *Nature* **395**, 194–198 (1998).

21. Philpott, A., Porro, E. B., Kirschner, M. W. & Tsai, L. H. The role of cyclin-dependent kinase 5 and a novel regulatory subunit in regulating muscle differentiation and patterning. *Genes Dev.* **11**, 1409–1421 (1997).
22. Patrick, G. N., Zhou, P., Kwon, Y. T., Howley, P. M. & Tsai, L. H. p35, the neuronal-specific activator of cyclin-dependent kinase 5 (Cdk5) is degraded by the ubiquitin-proteasome pathway. *J. Biol. Chem.* **273**, 24057–24064 (1998).
23. Imahori, K. *et al.* Possible role of tau protein kinases in pathogenesis of Alzheimer's disease. *Neurobiol. Aging* **19**, S93–98 (1998).
24. Baumann, K., Mandelkow, E. M., Biernat, J., Piwnicka-Worms, H. & Mandelkow, E. Abnormal Alzheimer-like phosphorylation of tau-protein by cyclin-dependent kinases Cdk2 and Cdk5. *FEBS Lett.* **336**, 417–424 (1993).
25. Michel, G. *et al.* Characterization of tau phosphorylation in glycogen synthase kinase-3 β and cyclin dependent kinase-5 activator (p23) transfected cells. *Biochim. Biophys. Acta* **1380**, 177–182 (1998).
26. Paudel, H. K. Phosphorylation by neuronal cdc2-like protein kinase promotes dimerization of Tau protein *in vitro*. *J. Biol. Chem.* **272**, 28328–28334 (1997).
27. Paudel, H. K., Lew, J., Ali, Z. & Wang, J. H. Brain proline-directed protein kinase phosphorylates tau on sites that are abnormally phosphorylated in tau associated with Alzheimer's paired helical filaments. *J. Biol. Chem.* **268**, 23512–23518 (1993).
28. Lovestone, S. *et al.* Alzheimer's disease-like phosphorylation of the microtubule-associated protein tau by glycogen synthase kinase-3 in transfected mammalian cells. *Curr. Biol.* **4**, 1077–1086 (1994).
29. Lovestone, S. & Reynolds, C. H. The phosphorylation of tau: a critical stage in neurodevelopment and neurodegenerative processes. *Neuroscience* **78**, 309–324 (1997).
30. Pant Veeranna, A. C., Pant, H. C. & Amin, N. Phosphorylation of human high molecular weight neurofilament protein (hNF-H) by neuronal cyclin-dependent kinase 5 (Cdk5). *Brain Res.* **765**, 259–266 (1997).
31. Sun, D., Leung, C. L. & Liem, R. K. H. Phosphorylation of the high molecular weight neurofilament protein (NF-H) by Cdk5 and p35. *J. Biol. Chem.* **271**, 14245–14251 (1996).
32. Guidato, S., Tsai, L. H., Woodgett, J. & Miller, C. C. Differential cellular phosphorylation of neurofilament heavy side-arms by glycogen synthase kinase-3 and cyclin-dependent kinase-5. *J. Neurochem.* **66**, 1698–1706 (1996).
33. Mandelkow, E. M. *et al.* Tau domains, phosphorylation, and interactions with microtubules. *Neurobiol. Aging* **16**, 355–362; discussion 362–353 (1995).
34. Lee, K. Y., Rosales, J. L., Tang, D. & Wang, J. H. Interaction of cyclin-dependent kinase 5 (Cdk5) and neuronal Cdk5 activator in bovine brain. *J. Biol. Chem.* **271**, 1538–1543 (1996).
35. Ahuja, H. S., Zhu, Y. & Zakeri, Z. Association of cyclin-dependent kinase 5 and its activator p35 with apoptotic cell death. *Dev. Genet.* **21**, 258–267 (1997).
36. Henchcliffe, C. & Burke, R. E. Increased expression of cyclin-dependent kinase 5 in induced apoptotic neuron death in rat substantia nigra. *Neurosci. Lett.* **230**, 41–44 (1997).
37. Shirvan, A. *et al.* Expression of cell cycle-related genes during neuronal apoptosis: is there a distinct pattern. *Neurochem. Res.* **23**, 767–777 (1998).
38. Zhang, Q., Ahuja, H. S., Zakeri, Z. F. & Wolgemuth, D. J. Cyclin-dependent kinase 5 is associated with apoptotic cell death during development and tissue remodeling. *Dev. Biol.* **183**, 222–233 (1997).
39. Matsubara, M. *et al.* Site-specific phosphorylation of synapsin I by mitogen-activated protein kinase and Cdk5 and its effects on physiological functions. *J. Biol. Chem.* **271**, 21108–21113 (1996).
40. Shuang, R. *et al.* Regulation of Munc-18/syntaxin 1A interaction by cyclin-dependent kinase 5 in nerve endings. *J. Biol. Chem.* **273**, 4957–4966 (1998).
41. Tsai, L. H., Takahashi, T., Caviness, V. S. Jr & Harlow, E. Activity and expression pattern of cyclin-dependent kinase 5 in the embryonic mouse nervous system. *Development* **119**, 1029–1040 (1993).
42. Binder, L. I., Frankfurter, A. & Rebhun, L. I. The distribution of tau in the mammalian central nervous system. *J. Cell Biol.* **101**, 1371–1378 (1985).
43. Greenberg, S. G., Davies, P., Schein, J. D. & Binder, L. I. Hydrofluoric acid-treated tau PHF proteins display the same biochemical properties as normal tau. *J. Biol. Chem.* **267**, 564–569 (1992).
44. Wolozin, B. L., Pruchnicki, A., Dickson, D. W. & Davies, P. A neuronal antigen in the brains of Alzheimer patients. *Science* **232**, 648–650 (1986).
45. Bursztajn, S. *et al.* Overexpression in neurons of human presenilin-1 or a presenilin-1 familial Alzheimer disease mutant does not enhance apoptosis. *J. Neurosci.* **18**, 9790–9799 (1998).
46. Fix, A. S., Ross, J. F., Stitzel, S. R. & Switzer, R. C. Integrated evaluation of central nervous system lesions: stains for neurons, astrocytes, and microglia reveal the spatial and temporal features of MK-801-induced neuronal necrosis in the rat cerebral cortex. *Toxicol. Pathol.* **24**, 291–304 (1996).
47. Tanaka, T., Iqbal, K., Trenkner, E., Liu, D. J. & Grundke-Iqbal, I. Abnormally phosphorylated tau in SY5Y human neuroblastoma cells. *FEBS Lett.* **360**, 5–9 (1995).
48. Hockenbery, D., Nunez, G., Millman, C., Schreiber, R. D. & Korsmeyer, S. J. Bcl-2 is an inner mitochondrial membrane protein that blocks programmed cell death. *Nature* **348**, 334–336 (1990).
49. Xia, Z., Dickens, M., Raingeaud, J., Davis, R. J. & Greenberg, M. E. Opposing effects of ERK and JNK-p38 MAP kinases on apoptosis. *Science* **270**, 1326–1331 (1995).

Acknowledgements

We thank D. Auprin, P. Davies, X. He, R. Neve and J. P. Vonsattel for reagents; Y. Zhou for antibody preparations; D. Smith for help with microscopy; and M. Greenberg, P. Lu, Y. Shi, G. Gill, Y. T. Kwon, D. Smith, V. Tannoch and J. Volker for critical reading of this manuscript. This work was partially supported by NIH grants to L.-H.T. L.-H.T. is an assistant investigator of the Howard Hughes Medical Institute, a Rita Allen Foundation scholar and a recipient of an Ester A. and Joseph Klingenstein Fund award.

Correspondence and requests for materials should be addressed to L.-H.T. (e-mail: li-huei_tsai@hms.harvard.edu).

Crystal structure of a lectin-like natural killer cell receptor bound to its MHC class I ligand

José Tormo^{*†}, Kannan Natarajan[‡], David H. Margulies[‡] & Roy A. Mariuzza^{*}

^{*} Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, 9,600 Gudelsky Drive, Rockville, Maryland 20850, USA

[†] Institut de Biologia Molecular de Barcelona, CSIC, Jordi Girona 18-26, 08034 Barcelona, Spain

[‡] Molecular Biology Section, Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

Natural killer (NK) cell function is regulated by NK receptors that interact with MHC class I (MHC-I) molecules on target cells. The murine NK receptor Ly49A inhibits NK cell activity by interacting with H-2D^d through its C-type-lectin-like NK receptor domain. Here we report the crystal structure of the complex between the Ly49A NK receptor domain and unglycosylated H-2D^d. The Ly49A dimer interacts extensively with two H-2D^d molecules at distinct sites. At one interface, a single Ly49A subunit contacts one side of the MHC-I peptide-binding platform, presenting an open cavity towards the conserved glycosylation site on the H-2D^d α 2 domain. At a second, larger interface, the Ly49A dimer binds in a region overlapping the CD8-binding site. The smaller interface probably represents the interaction between Ly49A on the NK cell and MHC-I on the target cell, whereas the larger one suggests an interaction between Ly49A and MHC-I on the NK cell itself. Both Ly49A binding sites on MHC-I are spatially distinct from that of the T-cell receptor.

NK cells are a fundamental component of the innate immune system and have the intrinsic ability to recognize and destroy certain virally infected and tumour cells. Susceptibility to lysis of cells with abnormal MHC-I expression is, in part, due to the lack of inhibition of NK-cell-mediated cytotoxicity, and reconstitution of normal MHC-I levels restores inhibition of lysis^{1–4}. These observations are consistent with the ‘missing self’ hypothesis⁵, in which NK receptor engagement by MHC-I inhibits NK-cell-mediated lysis of class-I-expressing target cells (self), thereby directing the cytolytic activity of NK cells against pathogen-infected or tumour cells that have lost class I expression (missing self).

Evidence in support of this hypothesis has been obtained through the molecular cloning and functional characterization of inhibitory receptors specific for MHC-I^{1–3}. To date, two superfamilies of NK receptors, the immunoglobulin superfamily and the C-type-lectin-like superfamily, have been shown to regulate NK cell cytotoxicity. NK receptors of the immunoglobulin superfamily (KIRs, ILTs, gp49, PIRs, LAIR-1) are type I transmembrane glycoproteins containing one or more immunoglobulin-like extracellular domains, whereas members of the C-type-lectin-like receptor family (Ly49A through J, NKR-P1, CD94/NKG2, CD69) are homo- or heterodimeric type II transmembrane glycoproteins, with each chain containing a single, extracellular C-type-lectin-like domain^{1–3}. Both superfamilies include activating and inhibitory receptors, such that NK cell function is regulated by a dynamic balance between positive signalling receptors (resulting in target cell lysis) and negative signalling receptors (preventing lysis). A common feature of the inhibitory NK receptors is the presence of cytoplasmic immunoreceptor tyrosine-based inhibition motifs that transduce negative signals^{1–3}. Activating NK receptors are coupled to other transmembrane proteins that contain cytoplasmic immunoreceptor tyrosine-based activation motifs.

Crystal structures of several unbound NK receptors from the immunoglobulin superfamily (KIR2DL1, KIR2DL2 and KIR2DL3) have been determined^{6–8}. Their overall fold exhibits two tandem immunoglobulin-like domains positioned at an acute elbow angle which shows some variability between receptors. The putative binding site for their MHC-I ligand, HLA-C, is located on the

outer surface of the elbow and spans both domains. The structure of the human CD94 NK receptor domain (NKD), one of the subunits of the heterodimeric CD94/NKG2 NK cell inhibitory receptor, has also been solved, revealing a variation of the canonical C-type-lectin fold⁹. However, the structure of an NK receptor complexed with a specific MHC-I ligand has not yet been described.

Here we report the crystal structure of the Ly49A NKD homodimer in complex with H-2D^d at 2.3 Å resolution. This constitutes the first atomic structure of a complete lectin-like inhibitory NK receptor and reveals intriguing similarities and differences to *bona fide* animal lectins and the CD94 subunit of the CD94/NKG2 NK receptor. The Ly49A NKD/H-2D^d complex has two distinct interfaces between the interacting molecules. The binding at one site is consistent with previous data on the interaction between receptors on the NK cell and MHC molecules on the target cell (*trans* interaction). The second, more extensive binding site strongly suggests a *cis* interaction between Ly49A and class I molecules on the same NK cell. Both binding sites are distinct from that of the T-cell receptor (TCR)¹⁰, indicating the possibility of a coordinated interaction of MHC-I with both classes of receptors in some cases.

Overall structure in the crystal

We prepared soluble forms of the full extracellular region of Ly49A, encompassing the stalk (a region that connects the transmembrane domain with the NKD) and the NKD, and H-2D^d by *in vitro* refolding from recombinant material expressed in *Escherichia coli*^{11,12}. Soluble Ly49A and H-2D^d have been shown, by surface plasmon resonance methods, to bind specifically to each other with affinities in the micromolar range¹¹; however, we failed to obtain crystals of the complex using the full extracellular segment of Ly49A. Suspecting that the Ly49A stalk may be too flexible, we used limited proteolysis to identify a compact, globular domain more amenable to crystallization. Digestion with trypsin resulted in the elimination of 60 amino-terminal amino acids that comprise most of the stalk. The resulting domain (Ly49A NKD; residues 127–262) was similarly produced and purified; it retains specific binding to H-2D^d and to appropriate monoclonal antibodies (data not shown). When mixed at high protein concentrations (~10 mg ml⁻¹) in near-

Table 1 Data collection and refinement statistics

Data Processing	
Total observations to 2.3 Å	128,909
Unique reflections	33,230
Completeness	94.7 (80.7)
I/σ	15.1 (6.2)
R_{sym} (%)	5.3 (16.1)
Refinement	
Data range (Å)	20.0–2.3
No. of reflections used (working set)	31,015
No. of reflections in R_{free} set	1,971
No. of protein non-hydrogen atoms	5,182
No. of solvent atoms	364
R.m.s.d. bond lengths (Å)	0.006
R.m.s.d. bond angles (°)	1.37
R.m.s.d. B factors (Å ²)	1.85
R_{crs} (%)†	19.8
R_{free} (%)‡	23.8

Values in parentheses indicate the specific value in the highest resolution shell (2.35–2.30 Å).
 $R_{\text{sym}} = \sum |I_j - \langle I \rangle| / \sum I_j$, where I_j is the intensity of an individual reflection, and $\langle I \rangle$ is the average intensity of that reflection.

† $R_{\text{crs}} = \sum |F_o| - |F_c| / \sum |F_o|$, where F_c is the calculated structure factor.

‡ R_{free} is as for R_{crs} but calculated for 6% of randomly chosen reflections that were omitted from the refinement.

physiological conditions (Tris-buffered saline, pH 7.5), Ly49A NKD and H-2D^d readily associate without the aid of precipitants or other additives, giving rise to microcrystals or precipitates. Slight modification of the conditions yielded crystals that were suitable for X-ray diffraction studies (see Methods). These crystals belong to the space group $P2_12_12_1$ with $a = 79.0$ Å, $b = 97.0$ Å, $c = 99.3$ Å, and contain one Ly49A dimer and one H-2D^d molecule in the asymmetric unit. The structure of the Ly49A NKD/H-2D^d complex was determined by molecular replacement and refined to an R_{crs} of 19.8% and R_{free} of 23.8% for all data extending to 2.3 Å spacing (Table 1).

The overall structure of the Ly49A NKD/H-2D^d complex and a view of its packing in the crystal are shown in Fig. 1a. Although the crystallographic asymmetric unit contains an Ly49A dimer and a single H-2D^d molecule, Ly49A interacts extensively with two H-2D^d molecules related by the crystal symmetry (Fig. 1b, d). These interactions define two distinct interfaces which we refer to as site 1 and site 2. The packing of the Ly49A/MHC-1 complex is dominated by the alternation of these two interfaces along one of the crystal axes.

The Ly49A NKD fold

As predicted by amino-acid sequence similarity, Ly49A NKD adopts a fold similar to that of the carbohydrate-recognition domains (CRD) of C-type animal lectins¹³ and the NKD of the human NK receptor subunit CD94 (ref. 9; Fig. 2). For example, superposition of Ly49A onto rat mannose-binding protein¹⁴ (MBP) results in a 1.2 Å root mean square (r.m.s.) deviation for 85 Cα pairs, whereas superposition onto CD94 NKD gives an r.m.s. deviation of 1.1 Å for 86 Cα pairs. Ly49A NKD contains all the main secondary-structure elements found in other members of the C-type-lectin-like superfamily, including two antiparallel β-sheets and two α-helices. The β-strand β2 acts as a connection between the two β-sheets formed by strands β0, β1 and β5 in the lower part of the molecule (using the standard view for C-type-lectin-like folds as shown in Fig. 2), and β3 and β4 strands in the upper part. Strand β2 also forms a short β-hairpin with strand β2'. The upper part of the molecule is characterized by a long stretch, connecting strands β2' and β3, which lacks regular secondary structure. This region corresponds to the Ca²⁺- and ligand-binding sites of *bona fide* C-type lectins.

Ly49A differs from most C-type lectins in the position of helix α2, which is rotated by about 23°. The position of this helix in Ly49A is similar to that of the tunicate lectin TC14 (ref. 15) which, like Ly49A, forms homodimers. In contrast to the rest of the C-type

lectin-like family members described so far, the crystal structure of CD94 NKD lacks helix α2, despite the fact that members of the Ly49 and CD94 families show strong sequence similarity in this region^{9,13} (Fig. 3). The corresponding residues in CD94 adopt a loop conformation in the homodimer⁹, although the structure of CD94 when paired with an NKG2 subunit is not known. As in other C-type-lectin-like structures referred to as 'long-form CRDs', both NKDs have an additional β strand (β0) at their N termini, which in CD94 is involved in non-covalent dimerization. In Ly49A, however, this strand is longer, with six residues rather than the three found in CD94. Overall, if one disregards differences in helix α2, the Ly49A structure is closer to CD94 than to other members of the C-type-lectin-like superfamily.

There are four intrachain disulphide bonds in Ly49A NKD (Fig. 2), two of which (C167–C253 and C232–C245) are the characteristic invariant disulphides found in all C-type animal lectins (Fig. 3). The third disulphide bond, C145–C150, is at a location similar to the third invariant disulphide characteristic of long-form C-type lectins. In Ly49A, however, these two bonded cysteines are separated by only four residues and are located on contiguous β-strands, whereas in other C-type lectins they are separated by about 10 residues. The fourth disulphide, C163–C251, which links the N terminus of strand β5 to the first turn of helix α1, is unique to the Ly49 family of NK receptors. An additional cysteine at position 195, which is not conserved in all members of the Ly49 family (Fig. 3), is unpaired.

The loops that connect the secondary-structure elements constitute the regions where Ly49A NKD and the C-type animal lectins differ most. Ly49A NKD, like CD94 (ref. 9), appears to lack the conserved Ca²⁺-binding site found in the Ca²⁺-dependent lectins of this structural family¹³. In Ly49A, the putative Ca²⁺-dependent carbohydrate-binding loop (residues 224–231 preceding strand β3) is shorter and displays a different conformation (Fig. 2). In the C-type lectins, in addition to its direct role in carbohydrate interaction, Ca²⁺ is integral to the structure and stabilizes the surrounding loops that are essential for ligand binding¹⁶. Some of the modifications present in Ly49A appear to ensure structural stability in the absence of metal ion. Of the characteristic loops of MBPs, loop L1 has been replaced by a β-hairpin (strands β2 and β2'), the equivalent to loop L3 is missing, and loop L4 seems more structured. The conserved N226 forms a Asn-pseudoturn, which is further stabilized by two hydrogen bonds between main-chain atoms of the loop and the side chain of N242. The position of this loop is determined by the packing of I227 against Y201 on strand β2. This small hydrophobic cluster replaces a second Ca²⁺-binding site that is conserved in MBPs. Despite these differences, residues D229, D241 and N242 are in a position similar to that of Ca²⁺-coordinating residues in C-type lectins (Fig. 3). Some of these are instead involved in interactions with H-2D^d at both site 1 and site 2 (see below).

The Ly49A homodimer

At the cell surface, Ly49A is a disulphide-linked homodimer^{1–3}. It is presumed that the stalk regions of the Ly49A subunits associate through a parallel coiled-coil. Secondary-structure predictions (data not shown) suggest, however, that the coiled-coil regions may not extend through the whole stalk, but instead may be interspersed with flexible portions that would confer mobility to the ectodomain. Size-exclusion chromatography and equilibrium ultracentrifugation have shown that, in solution, the Ly49A extracellular portion tends to dimerize non-covalently with a monomer/dimer equilibrium constant in the micromolar range¹¹. The two Ly49A NKD copies in the crystal asymmetric unit form a dimer, although their association diverges notably from strict twofold symmetry. Instead, they are related by a 161.3° rotation, and have significant local differences at some exposed loops and around helix α2. When the two Ly49A subunits are superimposed using 54

residues forming the core of the domain, they have an r.m.s. deviation of only 0.28 Å for their main-chain atoms. However, this value increases to 0.68 Å when all residues of the NKD are used in the superposition, indicating a degree of independent flexibility of the two subunits. The interface between the two subunits is relatively flat and buries a surface area of 940 Å². The two monomers form hydrogen bonds through their respective first β-strand (β0, residues 141–146), creating an extended six-stranded antiparallel sheet. Beneath the β-sheet, the subunits associate through a central hydrophobic region formed by juxtaposition of the carboxy-terminal half of helix α2 and an adjacent loop (residues 185–192). Asymmetric binding of the Ly49A subunits to two different H-2D^d molecules, plasticity at the hydrophobic dimer interface, and

absence of the stalk regions in the crystallized molecules probably contribute to the deviation from the perfect twofold dyad observed in the Ly49A dimer.

The dimer arrangement in Ly49A is very similar to that observed for the CD94 crystallographic dimer⁹ even though helix α2 has been replaced by a loop in the latter. In the tunicate lectin TC14, which forms a physiologically relevant dimer, the two subunits associate similarly through an intermolecular extended β-sheet and hydrophobic interactions¹⁵. However, in TC14, which lacks strand β0, like other short-form C-type lectins, antiparallel pairing between subunits is through strand β1 and helices α2 associate side by side, giving rise to an overall orientation for the dimer that is considerably different from that of the Ly49A and CD94 NKDs.

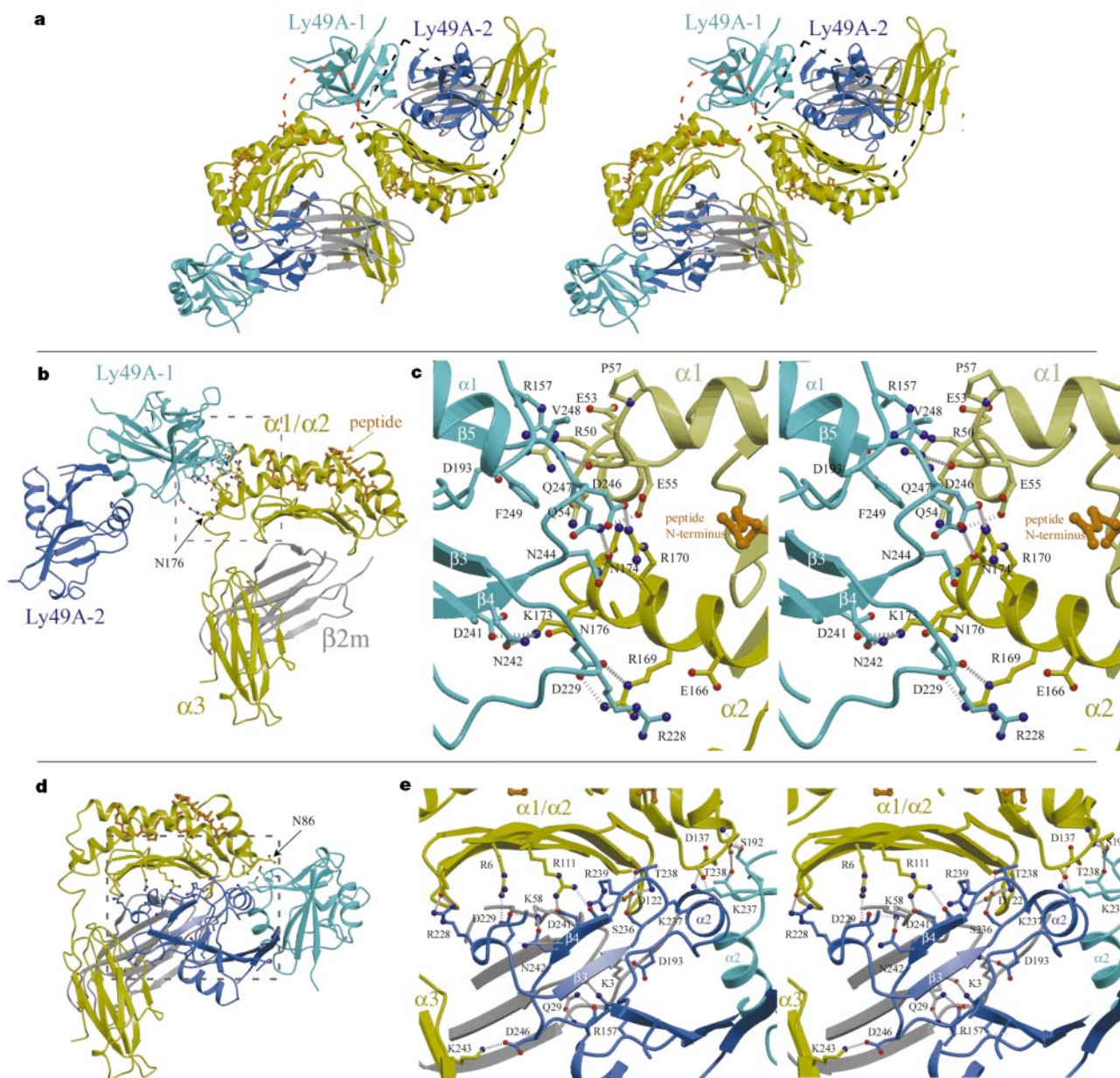


Figure 1 Structure of the Ly49A/H-2D^d complex and the interaction sites. **a**, Stereo diagram of the complex and crystal packing interactions. View shows the asymmetric interaction between an Ly49A dimer and two H-2D^d molecules related by the crystal symmetry. The H-2D^d heavy chain is yellow, β2m is grey, and the two Ly49A subunits, Ly49A-1 and Ly49A-2, are cyan and blue, respectively, whereas the viral peptide is shown in orange ball-and-stick representation. The two interaction surfaces, site 1 and site 2, are indicated by a red circle and a black rectangle, respectively. **b, c**, Detailed view

of site 1 showing the interactions using a common orientation based on the 'standard' view for the MHC-I molecule. **c**, Close up of the interactions at site 1. The view has been rotated around the horizontal relative to **b** and depicts the region highlighted by a box. **d, e**, As in **b, c** but for site 2. The two domains of the MHC-I peptide-binding platform are shown in cream (α1) and dark yellow (α2). Hydrogen bonds between Ly49A and H-2D^d are depicted as broken lines. All figures were drawn using programs BOBSCRIPT⁴⁹ and Raster3D⁵⁰.

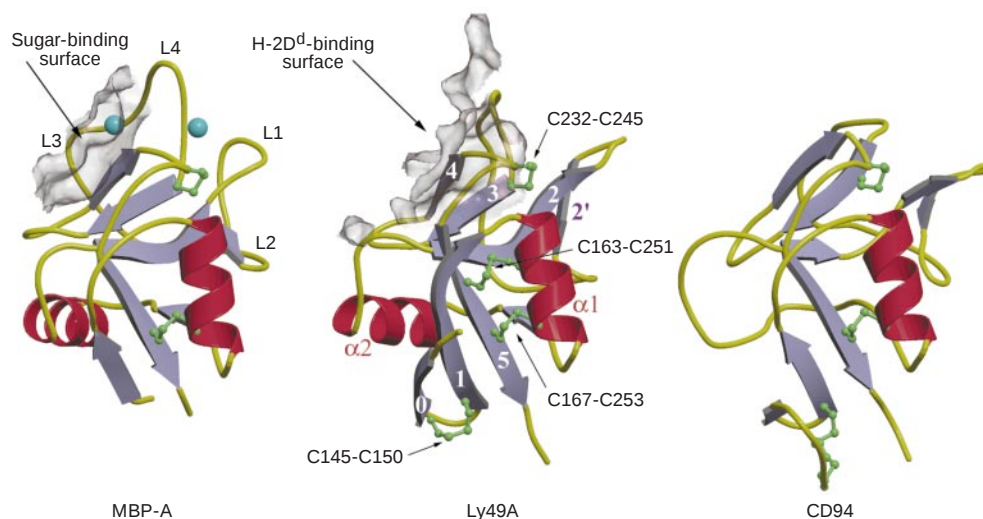


Figure 2 Structure of Ly49A NKD, and comparison with other members of the C-type lectin-like superfamily. Ribbon diagrams of the C-type lectin-like domains from Ly49A (centre), MBP-A (ref. 14) (left) and CD94 (ref. 9) (right) are shown in a common orientation obtained by pairwise superpositions. β -strands are blue and α -helices are red. The Ca^{2+} ions bound to MBP-A are shown as blue spheres, and the loop regions without regular secondary structure, three of them involved in Ca^{2+} coordination, are labelled L1 to L4. The disulphide bonds in Ly49A have been labelled using the numbering for MBP-A. Therefore, the first β -strand, which is not found in MBP-A, is labelled as $\beta 0$, and an additional strand, which forms a β -hairpin with strand $\beta 2$, is labelled as $\beta 2'$. Overall, Ly49A

has more similarity with CD94, especially in the unstructured loops linking the secondary-structure elements. However, whereas Ly49A has two α -helices like the rest of the C-type lectin-like domains, CD94 lacks helix $\alpha 2$. In Ly49A, helix $\alpha 2$ is rotated by about 23° , as compared with MBP-A and most other members of the C-type lectin-like superfamily. For MBP-A and Ly49A, the solvent-accessible surfaces in contact with their ligands, mannose and H-2D^d, respectively, are shown as semitransparent surfaces (calculated by the program VOLUMES; R. Esnouf, unpublished). For Ly49A, only residues contacting H-2D^d at site 1 were included in the calculations, with no modelled carbohydrate. Both proteins use analogous regions to contact their ligands, although the binding site on Ly49A is more extensive and does not involve a metal ion.

In Ly49A, the intersubunit extended β -sheet, which includes the N and C termini, fully exposes one of its faces to the solvent. Side chains exposed on this face are mainly hydrophobic or aromatic (V141, W143, M148, Y152, V154, L172 and the disulphide C145–C150), and are probably involved in the formation of a hydrophobic core at the stalk/NKD interface, as is the case in the MBP trimer^{13,17,18}.

In both the Ly49A and the CD94 structure, the apposition of the two monomers generates a relatively flat, continuous surface, which is located opposite to the N and C termini, and which includes the putative carbohydrate-contacting residues at opposite ends (Figs 1a and 2). In the Ly49A NKD/H-2D^d complex, this surface is involved, to different extents, in the interactions of the Ly49A subunits with two H-2D^d molecules. On the basis of sequence alignments and structural considerations, it has been proposed that the ligand-binding site of the human CD94–NKG2A heterodimer would also reside in that region⁹.

The Ly49A NKD/H-2D^d interfaces

In the crystal structure, Ly49A NKD contacts H-2D^d through two distinct interfaces. At site 1, which is less extensive, a single Ly49A subunit (designated Ly49A-1, see Fig. 1) binds at one side of the MHC-I peptide-binding platform but away from the peptide antigen (Fig. 1b, c). The interface on H-2D^d spans the N terminus of helix $\alpha 1$ and the C terminus of helix $\alpha 2$, and is located just above the conserved N-linked glycosylation site at position 176. The structural elements on Ly49A-1 involved in this interaction involve the loop preceding strand $\beta 3$, strand $\beta 4$ and the loop connecting it to strand $\beta 5$ (Figs 1c, 2 and 3). This binding site and the adjacent surface, which corresponds to the carbohydrate-binding site of MBPs, is roughly equivalent to the extended site of selectins^{13,19}, which are C-type lectins that mediate transient interactions between leukocytes and endothelial cells by binding complex carbohydrates¹³.

The total buried surface at site 1 is $994 \text{ \AA}^2$ (calculated using a probe radius of $1.4 \text{ \AA}$), with roughly equivalent contributions from

the $\alpha 1$ (45%) and $\alpha 2$ (55%) domains of H-2D^d. This is at the lower end of the average value of $1,600 (\pm 400) \text{ \AA}^2$ for protein–protein recognition sites²⁰, but is similar to the total buried surface in the heterophilic adhesion complex between human CD2 and CD58 (ref. 21). In addition, the buried surface at site 1, calculated for the present complex between unglycosylated proteins, does not include the potential contribution of carbohydrates at H-2D^d position 176 and, therefore, may underestimate the extent of the interaction between native glycosylated proteins. The site 1 interaction surface possesses a high degree of shape complementarity, as reflected by a shape correlation statistic (Sc)²² of 0.78. This is significantly greater than the Sc values for antigen–antibody complexes (0.64–0.68), and comparable to those for protease–protease inhibitor complexes (0.70–0.76), indicating a very precise match of the interacting surfaces²². Accordingly, water molecules are mainly excluded from the centre of the interface and are distributed at the periphery of the contact area.

Another feature of this interface is its highly hydrophilic nature, the interaction being dominated by charged interactions. The atoms buried at the interface are 35% non-polar, 22% neutral-polar and 43% charged, whereas an average protein–protein interface contains 56% non-polar, 29% neutral-polar and 15% charged atoms²⁰. Of the eight hydrogen bonds at this interface, five are formed between oppositely charged atoms (Table 2), which implies stronger interactions. Furthermore, there are two additional long-range ionic pairs: the side chains of H-2D^d residues E53 and E166 are located at short distances ($<4.8 \text{ \AA}$) from those of Ly49A residues R157 and R228, respectively. Thus, although the interacting surfaces are on average electrostatically neutral because of the intermingling of positively and negatively charged side chains on each molecule, there is an exquisite ionic complementarity between the interacting surfaces. These intimate electrostatic interactions can both ensure binding specificity and contribute to binding affinity. The polar interactions are distributed around a small patch of hydrophobic residues at the centre of the interface. This patch includes residues

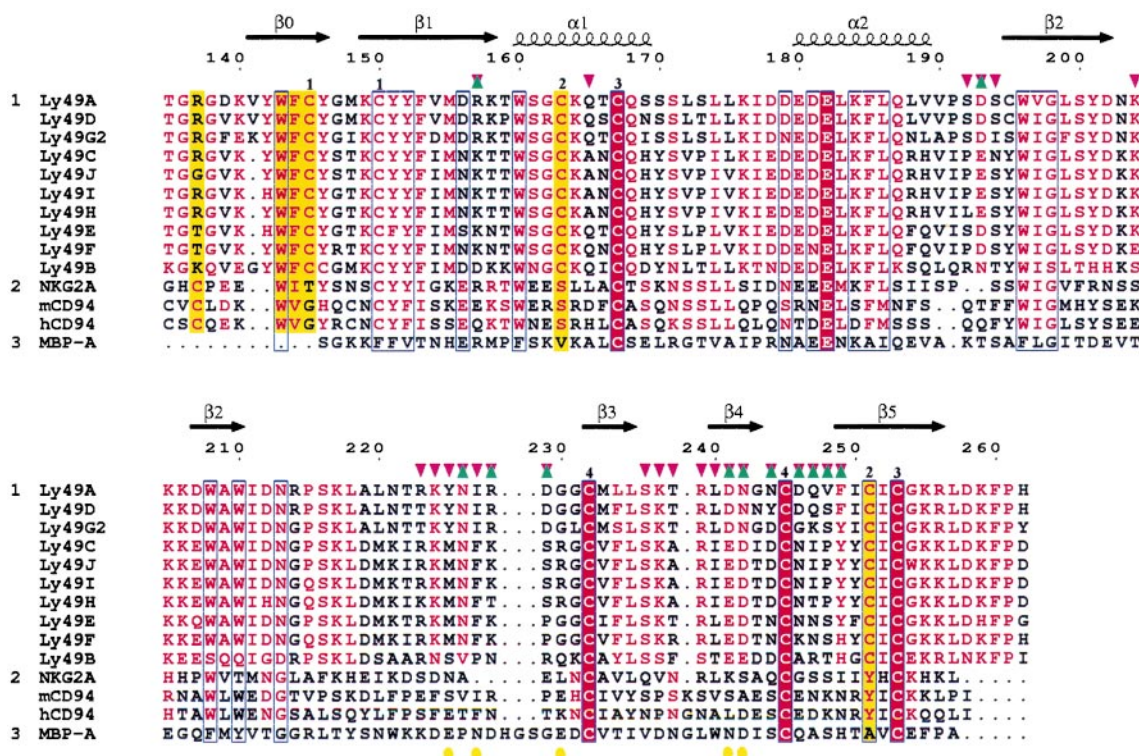


Figure 3 Structure-based sequence alignment of NKDs for selected NK cell receptors. NK cell receptors have been arranged into two groups. The first group contains 10 well-described members of the mouse Ly49 family, either inhibitory (A, B, C, E, F, G2, I and J) or activating (D and H). The second group includes the two subunits of the mouse inhibitory receptor CD94/NKG2A, and the CD94 subunit of the human receptor whose structure has been determined crystallographically⁹. The sequence of the CRD of the true C-type lectin MBP-A constitutes group 3. Residues strictly conserved are shown in white on a red background, residues well-conserved within a group (according to a Blosom62 matrix) in red, and the remainder in black. A blue frame denotes similarity across groups and a yellow box differences between conserved groups. The secondary-structure elements for Ly49A are denoted by squiggles (α -helices) and arrows (β -strands).

Symbols above the Ly49A sequence define molecular contacts (interatomic separations of less than 4.0 Å) with H-2D^d. A green triangle up denotes contacts at site 1, whereas those at site 2 are indicated by a purple triangle 'down'. The paired numbers (1–4) correspond to the bonded cysteines in Ly49A. Filled cyan ovals at the bottom indicate residues that coordinate the Ca²⁺ ion involved in carbohydrate-binding in MBP-A. Sequences were retrieved from GenBank and SwissProt databases (Ly49A, M25812; Ly49B, U10304; Ly49C, U56404; Ly49D, L78247; Ly49E, Q60652; Ly49F, Q60653; Ly49G2, S78689; Ly49H, U12889; Ly49I (ref. 37); Ly49J, AF110492; NKG2-A, AF109782; CD94, Q54708; human CD94, Q13241). The figure was drawn using ESPript (P. Gouet and F. Métoz, unpublished).

G56, P57 and the aliphatic portion of the side chain of Q54 at the N terminus of helix α 1 in H-2D^d, and V248, F249 on the loop connecting strands β 4 and β 5 in Ly49A. Residues Q54 of H-2D^d and V248 of Ly49A also form the only main chain/main chain hydrogen bond at this interface. As discussed below, we believe site 1 represents the known *trans* interaction between Ly49A on the NK cell and MHC class I on the target cell^{1–3,23}.

At site 2, which has a more extensive interface and buries 3,342 Å² of solvent-accessible surface, the Ly49A dimer is wedged into a cavity bounded by the heavy chain α 2 and α 3 domains of H-2D^d and β 2-microglobulin (β 2m). This cavity lies beneath the peptide-binding platform and partially overlaps the CD8-binding site^{24,25} (Figs 1d, e and 4b, c). The contribution of the Ly49A subunits (Ly49A-1 and Ly49A-2) is highly asymmetric, with 79% of the buried surface being contributed by Ly49A-2 (see Fig. 1). On the H-2D^d molecule, the three structural domains, α 1 α 2, α 3 and β 2m, contribute 60%, 15% and 25% to the interface, respectively. The flat surface of subunit Ly49A-2 lies against the surface of β 2m and the β -sheet that forms the bottom of the MHC-I peptide-binding platform, whereas the tip of the subunit approaches the H-2D^d α 3 domain (Fig. 1d). The binding surface at site 2 on Ly49A comprises, in addition to those involved in site 1, residues from the β 3– β 4 hairpin and the loop connecting helix α 2 to strand β 2. There is a minor contribution from the second subunit, Ly49A-1, which contacts the N terminus of helix α 2 and the C terminus of helix

α 1. This arrangement also locates the putative carbohydrate-binding site of Ly49A-1 in the proximity of the glycosylation site at N86 on the H-2D^d α 1 domain, which is conserved in all MHC-I molecules.

The interactions at site 2 are primarily polar and include the formation of 26 direct hydrogen bonds (Table 2), with many others bridged by numerous water molecules trapped at the interface. These solvent molecules appear to complement this otherwise poorly packed interface. The Sc value for this site, calculated without the inclusion of bound waters, is 0.54—much lower than the values obtained for antigen–antibody complexes (0.64–0.68)—and therefore reflects relatively poor shape complementarity²². It is within the range reported for TCR-peptide/MHC complexes (0.46–0.63)¹⁰, however, and comparable to the Sc value (0.58) for the CD2/CD58 complex²¹. Site 2 may represent a possible *cis* interaction between Ly49A and MHC-I on the same NK cell, as discussed below.

When comparing the structure of unbound H-2D^d (ref. 12) and its complex with Ly49A, the only substantial differences are the shifts in the positions of the α 3 and β 2m domains, which are equivalent to rotations of 25.7° and 6.4°, respectively. At the interfaces, conformational differences are mainly restricted to side-chain movements and are not larger than those observed between different H-2D^d crystal structures (refs 12 and 26; and J.T., K.N., D.H.M., R.A.M., unpublished data). Differences in the orientation of the α 3 and β 2m domains have also been noted in

Table 2 Interactions between Ly49A and H-2D^d

Site 1 Hydrogen bonds*		Site 2 Hydrogen bonds*	
H-2D ^d	Ly49A-1	H-2D ^d (heavy chain)	Ly49A-1
R50 N η 2	D193 O δ 1	Y85 O	T238 O γ
Q54 O	V248 N	N86 O	R239 N η 1
E55 O ϵ 2	Q247 N ϵ 2	D137 O δ 2	K237 N ζ
R169 N ϵ	D229 O δ 2	M138 N	S192 O γ
R169 N η 2	D229 O δ 1	A139 N	S192 O γ
R170 N η 2	D246 O δ 2		
K173 N ζ	D241 O δ 1	H-2D ^d (heavy chain)	Ly49A-2
N174 O δ	Q247 N ϵ 2		
Site 1 long-range ionic interactions†		S2 O	R228 N η 1
		S2 O	R228 N η 2
		R6 N η 1	D229 O δ 1
		R6 N η 2	D229 O δ 2
		R111 N η 2	L240 O
		R111 N η 1	D241 O δ 1
		D122 O δ 2	S236 O γ
		D122 O δ 2	T238 N
		D122 O δ 2	T238 O γ
		D122 O δ 1	R239 N η 2
		A136 O	D246 O δ 2
		K243 N ζ	Q165 N ϵ
		E227 O ϵ 1	Q247 N ϵ
		E232 O ϵ 2	
		H-2D ^d (β 2m)	Ly49A-2
		K3 N ζ	D193 O δ 1
		T4 O	R157 N η 1
		Q29 O	R157 N η 2
		Q29 O ϵ 1	V248 N
		Q29 N ϵ 2	V248 O
		K58 N ζ	D229 O δ 2
		K58 N ζ	N242 O δ

* Hydrogen bonds were calculated using standard parameters and a cut-off distance 2.5–3.4 Å.
† Long-range ionic interactions correspond to oppositely charged side chains with atoms located at distances shorter than 4.8 Å.

other MHC-I complexes. For example, a shift in the position of the HLA-A2 α 3 domain was observed in the structure of the CD8/HLA-A2 complex²⁴, whereas both the α 3 and β 2m domains of H-2K^b undergo a change in orientation in the 2C TCR/H-2K^b/dEV8 complex¹⁰. Whether or not these domain shifts in MHC-I are the consequence of engagement by NK receptor, CD8 or TCR, however, is unclear.

Basis of Ly49A binding and specificity

The various functional and binding data on the interaction of Ly49A and other members of the Ly49 family with MHC-I may now be reassessed in the context of the crystal structure of the Ly49A NKD/H-2D^d complex. Recognition by Ly49A requires that the H-2D^d molecule has bound a peptide, but, in striking contrast to the TCR, it is independent of the specific peptide sequence^{27,28}. On the basis of the analysis of Ly49A binding to various chimaeric MHC-I molecules and functional assays, the allelic specificity of Ly49A has been mapped to several regions on the α 1 α 2 domain of H-2D^d (refs 26, 29–32). Because Ly49A resembles true C-type lectins in sequence, the putative role of carbohydrates in the interaction has been studied extensively. Ly49A, like Ly49C, can bind to different sugars^{33,34}, especially sulphated polysaccharides, and the adhesion of these receptors to MHC-I ligands is sensitive to enzymatic or post-translational modification of the carbohydrate moiety on the MHC^{34,35}. Elimination of the glycosylation site at residue 176, but not residue 86, of H-2D^d considerably reduces the binding of Ly49A, although cells expressing this mutation were still partially resistant to killing³⁶; however, other studies have shown that recombinant unglycosylated MHC-I molecules are still recognized specifically by Ly49 receptors^{11,29,37}. These results suggest that, although carbohydrates attached to H-2D^d residue 176 seem to function as a part of the ligand structure for Ly49A and may modulate the affinity of the interaction, additional structural features are involved in recognition. It therefore appears that Ly49A, though not a true C-type lectin, reflects an evolutionary progression from a sugar-binding to a primarily protein-binding molecule that may retain sugar-binding capacity.

Binding of Ly49A to H-2D^d at site 1 is consistent with these data. By simultaneously contacting the two α -helices of the MHC-I α 1 α 2 domain at a site away from the bound peptide, recognition by Ly49A would clearly depend on the peptide-induced folding of this domain, but would be insensitive to the specific peptide sequence. Although one cannot discount long-range effects arising from particular peptide sequences, these seem improbable because this region shows little variation in the different MHC-I/peptide complexes analysed crystallographically³⁸. Interestingly, binding at site 1 would be consistent with a direct interaction between Ly49A and carbohydrates attached at H-2D^d N176. The binding interface

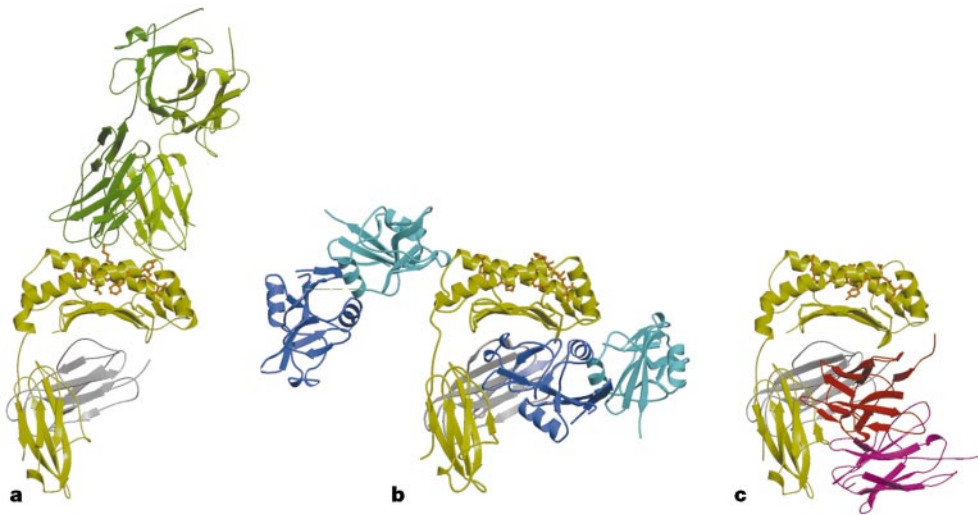


Figure 4 Binding of lymphocyte receptors to MHC-I. The complexes of mouse MHC-I molecules with the 2C TCR (ref. 10) (**a**), Ly49A (**b**) and CD8 $\alpha\alpha$ (ref. 25) (**c**) are depicted as ribbon diagrams in a common orientation based on the superposition of their respective MHC-I α 1 α 2 domains. The MHC-I heavy chains are yellow, β 2m is grey, and the peptides are shown in orange ball-and-stick representation. The α - and β -chains of the 2C TCR are in dark green and light green, respectively (**a**); the two Ly49A subunits, Ly49A-1 and Ly49-2, are in cyan and dark blue, respectively (**b**); and the two subunits of

the CD8 $\alpha\alpha$ homodimer, CD8 α -1 and CD8 α -2 are red and purple, respectively (**c**). For the Ly49A/H-2Dd complex, two Ly49A dimers interacting at the distinct binding sites are shown. The Ly49A dimer at site 1 binds to the polymorphic α -helices in a region away from the peptide and contiguous to the TCR binding-site. The Ly49A dimer at site 2 binds to a non-polymorphic region bounded by the heavy chain α 2 and α 3 domains and β 2m which partially overlaps the CD8-binding site.

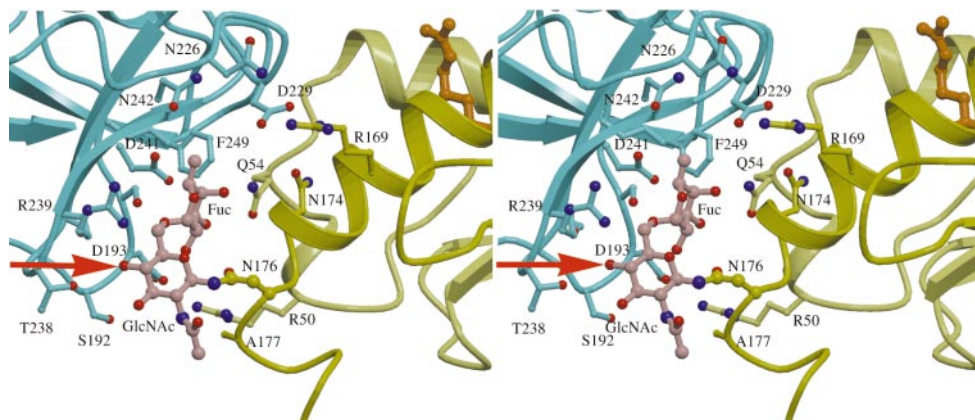


Figure 5 The putative carbohydrate-binding pocket at the Ly49A/H-2D^d interface. A close-up of the complex interface at site 1 is shown, centred at the open cavity around N176, a conserved glycosylation site in rodent MHC-I molecules. The orientation and the polypeptide chain representation are as in Fig. 1b. Two proximal carbohydrates, a GlcNAc and a fucose residue, have been modelled on the basis of crystal structures of MHC-I molecules deposited in the Protein Data Bank. The carbohydrate residues (pink) and

surrounding amino-acid side chains (cyan for Ly49A and yellow for H-2D^d) are shown in ball-and-stick representation; the branching fucose residue could fit well into the deep open pocket. The red arrow indicates position 4 in the GlcNAc residue where the rest of the oligosaccharide is attached, and which can establish further interactions along the flat surface of the Ly49A dimer (Fig. 1b).

presents an open cavity facing N176 that is surrounded by polar residues from both Ly49A and H-2D^d (Fig. 1b, c). Modelling of the carbohydrates attached to N176 suggests that this pocket can easily accommodate the proximal *N*-acetylglucosamine (GlcNAc) residue and, especially, a fucose residue branching from the first GlcNAc, a common modification of *N*-linked oligosaccharides of MHC-I molecules³⁸ (Fig. 5). The flat hydrophilic surface of Ly49A, contiguous to site 1, could also accommodate a larger number of carbohydrate residues; their post-translational modification, for example by sulphation, could have a role in modulating the Ly49A NKD/H-2D^d interaction³⁵. In addition, binding site 1 on H-2D^d contains various polymorphic residues (50, 55, 169, 173–174) that can, at least in part, explain the allelic specificity of Ly49A. These H-2D^d residues are matched by the high sequence variability of a contacting loop of Ly49A formed by residues 241–250 (Fig. 3). Site 1 is consistent with the expected *trans* orientation for NK receptor and MHC-I on opposite membranes, as the polypeptide termini linking Ly49A (a type II integral membrane protein) and H-2D^d to the membranes are pointing in opposite directions (Fig. 1b). With small rearrangements, site 1 is also compatible with one Ly49A dimer binding bivalently to two MHC-I molecules in *trans*. The subsequent crosslinking of Ly49A dimers by multivalently arrayed MHC-I on the target cell surface could lead to Ly49A-mediated signalling. This view is consistent with the available binding data obtained using purified components¹¹.

On the other hand, for site 2, there are no polymorphic H-2D^d residues at the interface that could account for the allelic specificity of Ly49A. As binding at this site requires $\alpha 1\alpha 2$, $\beta 2m$ and $\alpha 3$ in a defined relative orientation, specificity would need to be explained by polymorphic residues at distant positions affecting the segmental flexibility of the MHC-I molecule. In addition, the polypeptide termini of Ly49A and H-2D^d that link them to membranes are in the same orientation (Fig. 1c). Site 2 would be consistent with NK receptor and MHC-I being located on the same cell. In this respect, it should be noted that *cis* interactions between Ly49 and MHC-I on the NK cell itself have been proposed as a mechanism for the regulation of NK receptor expression and function^{39,40}. However, the stem portion of the molecule, extending from residues 60 to 127, is of sufficient length to permit a *trans* interaction through site 2 of the NKD and an MHC-I molecule on an opposite cell membrane.

Although most Ly49 family members probably interact with MHC-I similarly to Ly49A, data showing that MHC-I recognition

by Ly49C and Ly49I appears to depend not only on peptide occupancy, but also on its sequence^{37,41}, suggest that these two NK receptors may bind to sites on MHC-I that are at least partially distinct from those observed in the Ly49A/H-2D^d complex. For Ly49C, the sequence dependence has been mapped to the C-terminal part of the peptide (residue P7 of an octamer)⁴¹. Antibody blocking experiments further support the involvement of the C-terminal part of the MHC-I $\alpha 1$ helix in Ly49C binding to H-2D^b (ref. 37). Interestingly, CD94/NKG2 binding to HLA-E is also sensitive to substitutions at a C-terminal peptide position (P8 of a nonamer)⁴². These results suggest that a subgroup of the Ly49 family may bind to a distinct site involving the peptide C terminus that overlaps with the KIR binding site and perhaps the CD94/NKG2 site. In Fig. 3, in which the sequences in each group are arranged on the basis of similarity, Ly49C and Ly49I (the members sensitive to peptide sequence) appear to cluster together and have differences from other Ly49 members in the MHC-binding site. Thus, site 1 (and site 2) residues 157, 193, 226, 228, 229 and 241–249 are almost identical for Ly49C and Ly49I (and for Ly49J and Ly49H), but different from Ly49A.

Signalling

On the basis of the structure of the Ly49A NKD/H-2D^d complex, it is provocative to consider how MHC binding to Ly49A may contribute to the biochemical events required for Ly49A receptor redistribution, phosphorylation and transmission of inhibitory signals. The MHC-I on the target cell surface is multivalent and expressed at high density. Thus, the *trans* interaction of Ly49A with target cell MHC-I might be sufficient to induce the topological reorganization of the NK receptor and the subsequent accessibility to appropriate kinases, either by the interaction of inhibitory receptors with MHC-I alone, or by the concomitant interaction of inhibitory receptors in the vicinity of activating receptors. It is also conceivable that a simple *trans* interaction is not sufficient for full signalling through the inhibitory receptor, and that *cis* interactions may contribute to the modulation of the signalling events. Specific spatial orientation of Ly49A may result from simultaneous *cis* and *trans* interactions.

As shown in Fig. 4, the Ly49A-binding sites on MHC-I are spatially distinct from that of the TCR¹⁰. In agreement with the crystal structure, binding studies with recombinant Ly49A, H-2D^d and TCR¹¹, as well as functional studies in transgenic animals expressing a single chain H-2D^d (ref. 43), indicate that Ly49A is

capable of binding H-2D^d at the same time that the latter is complexed with a specific TCR. Such ligand sharing by two different receptors may constitute a means for regulating cytokine production by NK T cells, a specialized lymphocyte population expressing both NK receptors and TCRs specific for CD1/glycolipid complexes⁴⁴.

In summary, our results provide a firm basis for future analysis of the interaction between NK cell receptors and MHC-I. The complex has revealed two distinct binding sites for Ly49A on MHC-I, one of them unexpected. Structures of complexes of other members of the Ly49 family are needed to show whether these are common to all of them or whether certain members use different sites, as suggested by the dependence on peptide sequence for the binding of Ly49C and Ly49I. Furthermore, binding and crystallographic studies with naturally glycosylated MHC-I should fully unravel the role of carbohydrates in these interactions. □

Methods

Protein expression, refolding and purification

The full extracellular portion of Ly49 (amino acids 67–262) was expressed in *E. coli* as inclusion bodies, solubilized, refolded and purified as described¹¹. Limited proteolysis with trypsin resulted in the removal of ~60 residues from the N terminus. The resulting domain (referred to here as Ly49A NKD), including residues 127–262, was similarly expressed in *E. coli*, refolded and purified. Mouse β_2m and the soluble portion of the H-2D^d heavy chain were separately expressed as inclusion bodies in *E. coli*, solubilized, refolded in the presence of synthetic peptide P18-I10 (RGPGRFVFTI; derived from residues 318–327 of the human immunodeficiency virus IIIB gp120 envelope protein) and purified as described¹².

Crystallization and X-ray data collection

Crystals of the Ly49A NKD/H-2D^d complex were grown at 4 °C in hanging drops by mixing 1 μ l of protein solution (9 mg ml⁻¹ of Ly49A NKD and H-2D^d at a 1.5:1 molar ratio in 20 mM HEPES, 100 mM NaCl, pH 7.5) with an equal volume of the reservoir solution (9% polyethylene glycol monomethyl ether 550, 50 mM magnesium acetate, 100 mM MES, pH 6.5). The presence of both H-2D^d and Ly49A molecules, in a molar ratio of 1:2, was demonstrated by PAGE. For cryogenic data collection, crystals were harvested in a modified reservoir solution, transferred to harvest buffer containing 10% ethylene glycol and flash-cooled by plunging into liquid propane. A data set was collected at beam line ID14 of the Advanced Photon Source (Argonne National Laboratory) using a Quantum-4 CCD detector ($\lambda = 1.0$ Å). Crystals belong to the space group $P2_12_12_1$ with unit-cell dimensions $a = 79.0$ Å, $b = 97.0$ Å, $c = 99.3$ Å, and contain one H-2D^d molecule and one Ly49A dimer per asymmetric unit. The data were indexed, integrated and scaled with the HKL package⁴⁵ to yield a data set 94.7% complete at 2.3 Å ($R_{\text{sym}} = 5.3\%$; Table 1).

Structure determination

The structure was solved by molecular replacement using the program AMoRe⁴⁶. Search probes consisted of an H-2D^d model refined at 1.95 Å resolution (ref. 12; and J.T., K.N., D.H.M. and R.A.M., unpublished data) and the human CD94 monomer⁹ (Protein Data Bank accession code 1b6e) pruned of those structural elements that were suspected to differ between CD94 and Ly49A. Clear solutions were found in the cross-rotation, and translation functions for H-2D^d and only one of the Ly49A molecules. This partial solution was rigid-body refined using four domains ($\alpha 1\alpha 2$, $\alpha 3$, β_2m , Ly49A) and the resulting model was fixed during a new molecular replacement search for the second Ly49A copy. This search yielded correct solutions for the rotation–translation parameters for the three molecules in the asymmetric unit which when combined resulted in a R_{cryst} of 47.0% for data between 10.0 Å and 3.5 Å. After five-domain rigid-body refinement ($\alpha 1\alpha 2$, $\alpha 3$, β_2m and two Ly49A subunits), the R_{cryst} dropped to 45.7% ($R_{\text{free}} = 46.0\%$) for data between 10.0 Å and 3.0 Å. Electron density maps calculated on the basis of these phases showed unambiguous density for the second α -helix ($\alpha 2$) of the typical C-type domain fold and other structural elements absent in the search model, thus confirming the correctness of the molecular replacement solution.

Refinement

Crystallographic refinement was carried out with CNS⁴⁷ using standard protocols, which included a bulk solvent correction and overall anisotropic B -factor scaling. Refinement was alternated with manual rebuilding in the graphics program O⁴⁸. No non-crystallographic symmetry restraints were applied to the two Ly49A molecules because of their very different packing environments and conformational differences evident during their modelling. Solvent molecules were included at the final stages of refinement at significant positive electron densities (3.5σ in $F_o - F_c$ maps) and after hydrogen-bonding requirements. All parts of the complex are well ordered, with the exception of the CD8-binding loop in the $\alpha 3$ domain of H-2D^d and a few residues in the N and C termini of Ly49A. The current refined atomic model of the complex comprises residues 2–275 of the H-2D^d heavy chain, residues 1–99 of β_2m , the decamer antigenic peptide, residues 126–134 and 140–258 from the Ly49A-1 subunit, residues 139–259 from the Ly49A-2 subunit, and 268 solvent atoms. For the non-glycine residues, 88% of the main-chain torsion angles lie in

the most favoured regions of the Ramachandran plot and none in the disallowed regions. The present R_{cryst} is 19.8%, and R_{free} is 23.8% for all data ($F > 0$) between 20 Å and 2.3 Å.

Received 18 August; accepted 20 October 1999.

1. Lanier, L. L. NK cell receptors. *Annu. Rev. Immunol.* **16**, 359–393 (1998).
2. Yokoyama, W. M. in *Fundamental Immunology* (ed. Paul, W. E.) 575–603 (Lippincott-Raven, Philadelphia, 1999).
3. Long, E. O. Regulation of immune responses through inhibitory receptors. *Annu. Rev. Immunol.* **17**, 875–904 (1999).
4. Kärre, K., Ljunggren, H. G., Piontek, G. & Kiessling, R. Selective rejection of H-2-deficient lymphoma variants suggests alternative immune defense strategy. *Nature* **319**, 675–678 (1986).
5. Ljunggren, H. G. & Kärre, K. In search of the “missing self”: MHC molecules and NK cell recognition. *Immunol. Today* **11**, 237–244 (1990).
6. Fan, Q. R. *et al.* Structure of the inhibitory receptor for human natural killer cells resembles haematopoietic receptors. *Nature* **389**, 96–100 (1997).
7. Snyder, G. A., Brooks, A. G. & Sun, P. D. Crystal structure of the HLA-Cw3 allotype-specific killer cell inhibitory receptor KIR2DL2. *Proc. Natl Acad. Sci. USA* **96**, 3864–3869 (1999).
8. Maenaka, K., Juji, T., Stuart, D. I. & Jones, E. Y. Crystal structure of the human p58 killer cell inhibitory receptor (KIR2DL3) specific for HLA-Cw3-related MHC class I. *Structure* **7**, 391–398 (1999).
9. Boyington, J. C. *et al.* Structure of CD94 reveals a novel C-type lectin fold: implications for the NK cell-associated CD94/NKG2 receptors. *Immunity* **10**, 75–82 (1999).
10. Garcia, K. C., Teyton, L. & Wilson, I. A. Structural basis of T cell recognition. *Annu. Rev. Immunol.* **17**, 369–397 (1999).
11. Natarajan, K. *et al.* Interaction of the NK cell inhibitory receptor Ly49A with H-2D^d: identification of a site distinct from the TCR site. *Immunity* **11**, 591–601 (1999).
12. Li, H., Natarajan, K., Malchiodi, E. L., Margulies, D. H. & Mariuzza, R. A. Three-dimensional structure of H-2D^d complexed with an immunodominant peptide from human immunodeficiency virus envelope glycoprotein 120. *J. Mol. Biol.* **283**, 179–191 (1998).
13. Weis, W. I., Taylor, M. E. & Drickamer, K. The C-type lectin superfamily in the immune system. *Immunol. Rev.* **163**, 19–34 (1998).
14. Weis, W. I., Kahn, R., Fourme, R., Drickamer, K. & Hendrickson, W. A. Structure of the calcium-dependent lectin domain from a rat mannose-binding protein determined by MAD phasing. *Science* **254**, 1608–1615 (1991).
15. Poget, S. F. *et al.* The structure of a tunicate C-type lectin from *Polyandrocarpa misakiensis* complexed with D-galactose. *J. Mol. Biol.* **290**, 867–879 (1999).
16. Ng, K. K.-S., Park-Snyder, S. & Weis, W. I. Ca²⁺-dependent structural changes in C-type mannose-binding proteins. *Biochemistry* **37**, 17965–17976 (1998).
17. Weis, W. I. & Drickamer, K. Trimeric structure of a C-type mannose-binding protein. *Structure* **2**, 1227–1240 (1994).
18. Sheriff, S., Chan, C. Y. & Ezekowitz, R. A. B. Human mannose-binding protein carbohydrate recognition domain trimerizes through a triple α -helical coiled-coil. *Nature Struct. Biol.* **1**, 789–794 (1994).
19. Graves, B. J. *et al.* Insight into E-selectin/ligand interaction from the crystal structure and mutagenesis of the lec/EGF domains. *Nature* **367**, 532–538 (1994).
20. Lo Conte, L., Chothia, C. & Janin, J. The atomic structure of protein–protein recognition sites. *J. Mol. Biol.* **285**, 2177–2198 (1999).
21. Wang, J.-H. *et al.* Structure of a heterophilic adhesion complex between the human CD2 and CD58 (LFA-3) counterreceptors. *Cell* **97**, 791–803 (1999).
22. Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950 (1993).
23. Karlhofer, F. M., Ribaldo, R. K. & Yokoyama, W. M. MHC class I alloantigen specificity of Ly-49⁺ IL-2 activated natural killer cells. *Nature* **358**, 66–70 (1992).
24. Gao, G. F. *et al.* Crystal structure of the complex between human CD8 $\alpha\alpha$ and HLA-A2. *Nature* **387**, 630–634 (1997).
25. Kern, P. S. *et al.* Structural basis of CD8 coreceptor function revealed by crystallographic analysis of a murine CD8 $\alpha\alpha$ ectodomain fragment in complex with H-2K^b. *Immunity* **9**, 519–530 (1998).
26. Achour, A. *et al.* The crystal structure of H-2D^b MHC class I complexed with the HIV-1-derived peptide P18-I10 at 2.4 Å resolution: implications for T cell and NK cell recognition. *Immunity* **9**, 199–208 (1998).
27. Correa, I. & Raulat, D. H. Binding of diverse peptides to MHC class I molecules inhibits target cell lysis by activated natural killer cells. *Immunity* **2**, 61–71 (1995).
28. Orihuela, M., Margulies, D. H. & Yokoyama, W. M. The natural killer cell receptor Ly49A recognizes a peptide-induced conformational determinant on its major histocompatibility complex class I ligand. *Proc. Natl Acad. Sci. USA* **93**, 11792–11797 (1996).
29. Matsumoto, N., Ribaldo, R. K., Abastado, J. P., Margulies, D. H. & Yokoyama, W. M. The lectin-like NK cell receptor Ly-49A recognizes a carbohydrate-independent epitope on its MHC class I ligand. *Immunity* **8**, 245–254 (1998).
30. Sundbäck, J. *et al.* The $\alpha 2$ domain of H-2D^d restricts the allelic specificity of the murine NK cell inhibitory receptor Ly-49A. *J. Immunol.* **160**, 5971–5978 (1998).
31. Karlhofer, F. M., Ribaldo, R. K. & Yokoyama, W. M. MHC class I alloantigen specificity of Ly-49⁺ IL-2 activated natural killer cells. *Nature* **358**, 66–70 (1992).
32. Brennan, J., Mahon, G., Mager, D. L., Jefferies, W. A. & Takei, F. Recognition of class I major histocompatibility complex molecules by Ly-49: specificities and domain interactions. *J. Exp. Med.* **183**, 1553–1559 (1996).
33. Daniels, B. F., Nakamura, M. C., Rosen, S. D., Yokoyama, W. M. & Seaman, W. E. Ly49A, a receptor for H-2D^d, has a functional carbohydrate recognition domain. *Immunity* **1**, 785–792 (1994).
34. Brennan, J., Takei, F., Wong, S. & Mager, D. L. Carbohydrate recognition by a natural killer cell receptor, Ly-49C. *J. Biol. Chem.* **270**, 9691–9694 (1995).
35. Chang, C. S. & Kane, K. P. Evidence for sulphate modification of H-2D^d on N-linked carbohydrate(s): possible involvement in Ly-49A interaction. *J. Immunol.* **160**, 4367–4374 (1998).
36. Lian, R. H., Freeman, J. D., Mager, D. L. & Takei, F. Role of conserved glycosylation site unique to murine class I MHC in recognition by Ly-49 NK cell receptor. *J. Immunol.* **161**, 2031–2036 (1998).
37. Hanke, T. *et al.* Direct assessment of MHC class I binding by seven Ly49 inhibitory NK cell receptors. *Immunity* **11**, 67–77 (1999).

38. Madden, D. R. The three-dimensional structure of peptide-MHC complexes. *Annu. Rev. Immunol.* **13**, 587–622 (1995).
39. Käse, A., Johansson, M. H., Olsson-Alheim, M. Y., Kärre, K. & Höglund, P. External and internal calibration of the MHC class I-specific receptor Ly49A on murine natural killer cells. *J. Immunol.* **161**, 6133–6138 (1998).
40. Andersson, M. *et al.* MHC class I mosaic mice reveal insights into control of Ly49C inhibitory receptor expression in NK cells. *J. Immunol.* **161**, 6475–6479 (1998).
41. Franksson, L. *et al.* Peptide dependency and selectivity of the NK cell inhibitory receptor Ly-49C. *Eur. J. Immunol.* **29**, 2748–2758 (1999).
42. Valés-Gómez, M., Reyburn, H. T., Erskine, R. A., López-Botet, M. & Strominger, J. L. Kinetics and peptide dependency on the binding of the inhibitory NK receptor CD94/NKG2-A and the activating receptor CD94/NKG2-C to HLA-E. *EMBO J.* **18**, 4250–4260 (1999).
43. Chung, D. H. *et al.* NK and CTL recognition of a single chain H-2D^d molecule: distinct sites of H-2D^d interact with NK and T cell receptors. *J. Immunol.* **163**, 3699–3708 (1999).
44. Bendelac, A., Rivera, M. N., Park, S.-H. & Roark, J. H. Mouse CD1-specific NK1 T cells: development, specificity, and function. *Annu. Rev. Immunol.* **15**, 535–562 (1997).
45. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–327 (1997).
46. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
47. Brünger, A. T. *et al.* Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
48. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building models in electron density maps and the location of errors in those models. *Acta Crystallogr. A* **47**, 110–119 (1991).
49. Esnouf, R. M. An extensively modified version of Molscrip that includes greatly enhanced colouring capabilities. *J. Mol. Graph.* **15**, 133–138 (1997).
50. Merrit, E. A. & Bacon, D. J. Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).

Acknowledgements

We thank W. Yang and B. A. Fields for their help with data collection, L. Boyd and R. Carey for assistance with protein expression and characterization, M. Garfield for protein sequencing, and J. C. Boyington and P. D. Sun for providing coordinates of the CD94 structure before release. We thank W. Yokoyama for encouragement and for critical comments on the manuscript. We also thank the staff at the Advanced Photon Source, Argonne National Laboratory, which is operated by the Department of Energy, Office of Basic Energy Sciences. This work was supported, in part, by grants from the National Institutes of Health and the National Multiple Sclerosis Society (R.A.M.) and the Spanish “Comision Interministerial de Ciencia y Tecnologia” (J.T.).

Correspondence and requests for materials should be addressed to R.A.M. (e-mail: mariuzza@carb.nist.gov). or D.H.M. (e-mail: dhm@nih.gov). Atomic coordinates have been deposited in the Protein Data Bank under accession number 1qo3.

Early planet formation as a trigger for further planet formation

Philip J. Armitage* & Brad M. S. Hansen

Canadian Institute for Theoretical Astrophysics, University of Toronto, Toronto, Ontario M5S 3H8, Canada

Recent discoveries of extrasolar giant planets at small orbital radii^{1,2}, or having significant orbital eccentricities, suggest that the planets interacted with the disks of dust and gas from which they and the central stars formed^{3–6}. Here we show that if a gas-giant planet reaches a mass of 4–5 jovian masses sufficiently early, when the protoplanetary disk is still massive, an otherwise stable disk will fragment into additional planetary bodies. This process of catastrophic planet formation could account for the apparent difference¹ in the distribution of the masses of massive planets and brown dwarfs around other stars, and the existence of young stars that appear to have dissipated their disks at a very early age⁷. Subsequent gravitational interactions^{5,6,8,9} between the first planet to form and the additional planets could lead to planetary systems comprising a small number of massive planets in eccentric orbits.

The planet–disk interaction has been studied extensively for low-mass protoplanetary disks^{3,10–14}. This is appropriate for the protosolar nebula where the rate-limiting step, that is, the assembly of the cores of the giant planets from smaller bodies¹⁵, is believed to require timescales comparable to the lifetimes of protoplanetary disks, which are observed⁷ to last $\sim 10^6$ – 10^7 years. However, some extrasolar giant planets could form more rapidly, either through direct hydrodynamic collapse¹⁶, or through accelerated core formation in disks that are significantly more massive¹⁷ than the minimum-mass solar nebula¹⁸. More broadly, if angular-momentum transport mechanisms other than self-gravity are inefficient in disks where the ionization fraction is low (no purely hydrodynamic instabilities that lead to outward angular-momentum transport are known to exist in keplerian disk flows¹⁹), then the outer regions of protoplanetary disks may remain only marginally stable against gravitational instability even at late evolutionary epochs^{20,21}. In either case, planet–disk interactions could occur while the effects of disk self-gravity are still significant.

The local stability of a gaseous disk against gravitational instability depends upon the balance between thermal pressure and self-gravity. At a point in a disk with sound speed c_s , surface density Σ , and angular velocity Ω , the controlling parameter²² is Toomre's Q , defined as:

$$Q = \frac{c_s \Omega}{\pi G \Sigma}$$

Here G is the gravitational constant. Numerical simulations^{23,24} show that non-axisymmetric instabilities set in at $Q \lesssim 1.5$, and become increasingly violent for smaller values of $Q \approx 1$. We consider the relatively cool outer regions of the disk, at radii of several astronomical units, and set our initial conditions such that the disk is marginally unstable, and has properties comparable to the upper end of the mass distribution of T Tauri disks inferred from millimetre-wavelength observations¹⁷, which have disk masses $m_{\text{disk}} \approx 0.1 m_{\odot}$, where $m_{\odot}$ is the solar mass. This is about an order of magnitude greater than the canonical minimum mass solar nebula value of $\sim 10^{-2} m_{\odot}$, although even for our Solar System the

initial disk mass may have been substantially in excess of this minimum^{25,26}. The surface density profile is taken to be:

$$\Sigma = \Sigma_0 r^{-3/2} \left(1 - \sqrt{\frac{r_{\text{in}}}{r}} \right)$$

Here r_{in} is the inner edge of the simulated disk annulus. Σ_0 and c_s are chosen such that $m_{\text{disk}} = 0.1$, and the ratio of disk scale height h to radius r at the outer edge is $(h/r) = 0.075$. With these parameters $Q > 1.5$ at all radii, and the disk is everywhere close to stable against gravitational instability, as expected if it is the endpoint of an earlier phase of violent gravitational instabilities that drive rapid angular-momentum transport²¹.

Figure 1 shows the evolution of the disk, computed using a lagrangian hydrodynamics code. The isolated disk is shown at time $t = 512$, in units where $\Delta t = 1$ corresponds to the dynamical time, Ω^{-1} , at r_{in} . This run shows weak spiral arms in the outer disk, as expected from the Q profile on the basis of previous simulations of gravitationally unstable disks^{23,24}, and is amply stable against fragmentation. The disk surface density does not evolve significantly over this relatively short interval, as expected for a thin disk where the efficiency of angular-momentum transport from gravitational instabilities, if parametrized approximately using an equivalent Shakura–Sunyaev α parameter²⁷, corresponds to a fairly small effective $\alpha \approx 10^{-2}$.

We now consider the evolution of the same star–disk system with an embedded planet of initial mass $m_p = 10^{-3} m_*$, where m_* is the mass of the star. For seeds of this mass and smaller, the additional potential fluctuations induced by the planet at the Lindblad resonances¹² are small compared to the background fluctuations due to the disk's own self-gravity, as measured in the control run. This is shown in Fig. 2. Neither the mass resolution nor the equation of state are realistic enough to model the internal structure of individual planets, so we focus solely on their influence on the disk, for which purpose details of their internal structure are unimportant.

The presence of a Jupiter mass planet significantly modifies the disk evolution. A partial gap is cleared in the disk on the dynamical timescale at r_p , bounded by a strongly compressed gravitational wake attached to the planet. This forms part of an $m = 2$ pattern (where m is the azimuthal mode number of the perturbation) of strong spiral arms, along with weaker transient spiral features excited in the disk by the combination of gravitational instability and planetary perturbation. The presence of a gap fails to prevent ongoing accretion along the spiral arms at a rate of accretion $\dot{m}_p \propto m_p$, with the planet mass increasing by a factor e on a timescale of a few planetary orbits. Continuing accretion is expected since the disk viscosity needs to be significantly lower than the values expected in a gravitationally unstable disk in order to inhibit accretion altogether¹¹. Much of this mass accumulates in a resolved, strongly tidally distorted disk surrounding the planet. As the planet mass grows, the overdensity in the spiral arms and at the Lindblad resonances increases while the background surface density profile is unable to evolve on as rapid a timescale. This essential imbalance in timescales is expected to be valid even for the formation of giant planets in a minimum-mass solar nebula²⁶, and is therefore a robust prediction for the more massive disks studied here. The rapid growth in mass leads to an increased amplitude of potential fluctuations, as shown in Fig. 2, decreasing disk stability, and inevitable fragmentation at the gap edges, shown in the lower panels of Fig. 1. For these disk parameters and equation of state this occurs at $m_p = 4 \times 10^{-3}$ – $5 \times 10^{-3} m_*$. Once this mass is reached, rapid fragmentation into numerous planetary mass bodies occurs near both the inner and outer Lindblad resonances.

Several additional calculations were used to check the sensitivity of the results to the initial conditions and numerical method. With the same initial conditions, fragmentation occurs at the same final planet mass in lower resolution simulations with 10,000 and 20,000 particles, although the initial masses of fragments do vary. In all

* Present address: Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Strasse 1, D-85740 Garching, Germany.

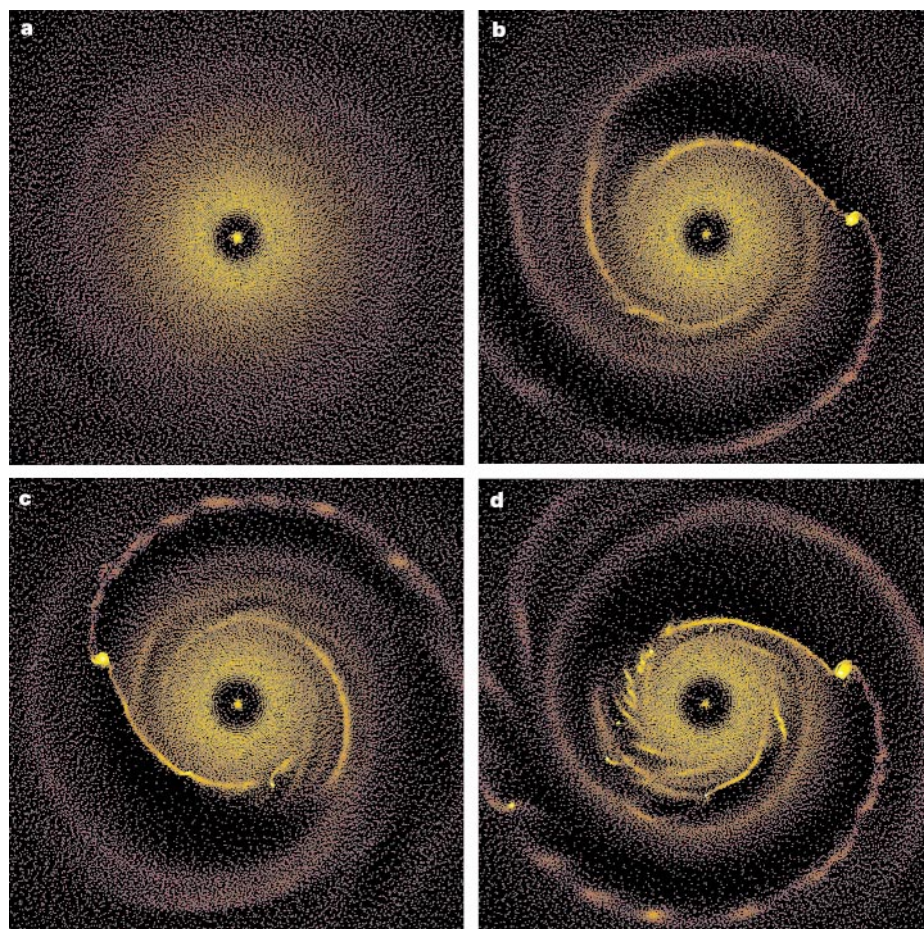


Figure 1 The evolution of the disk surface density, with and without an embedded massive planet. We compute the disk structure using a smooth particle hydrodynamics (SPH) code^{28,29}, with individual time steps for the particles and a tree structure for computing the gravitational forces³⁰. We use 60,000 SPH particles, an isothermal equation of state, and standard artificial viscosity parameters²⁹. We have verified that torques due to gravity and pressure forces dominate over those attributed to artificial viscosity in determining the planet–disk evolution. Additional support for collapsed objects is provided by limiting the force resolution by using a minimum SPH smoothing length, $h_{\min} = 0.075$. The central star is treated as a smoothed point mass chosen to give a keplerian potential at $r > r_{\text{in}}$. We use units in which $r_{\text{in}} = m_{\star} = 1$, and set the outer disk edge at $r_{\text{out}} = 25$. **a**, The isolated disk evolved to a time $t = 512$; **b**, the same disk at the same time but with a planet, initially of $10^{-3} m_{\star}$, orbiting in a coplanar circular orbit at

$r_p = 12.5$. The planet is treated as a point mass smoothed on a scale $h_p = 0.1$. The colours signify density on a logarithmic scale. In the presence of a planet the spiral structure is significantly amplified, and a partial gap has been cleared in the disk material. **c**, By $t = 608$, when the planetary mass (plus surrounding circumplanetary disk) has reached $4 \times 10^{-3} - 5 \times 10^{-3} m_{\star}$, the disk near the inner Lindblad resonance has become unstable and fragments into additional planets. Thereafter rapid destruction of the disk occurs. **d**, By $t = 736$ numerous fragments have formed, including one near the outer Lindblad resonance, and are accreting rapidly. t , time such that $\Delta t = 1$ for dynamical time Ω^{-1} at r_{in} ; r , disk radius; r_{in} , inner edge of disk annulus; r_{out} , outer edge of disk annulus; $m_{\star}$, mass of star; h_p , gravitational softening length of planet; r_p , planetary radius.

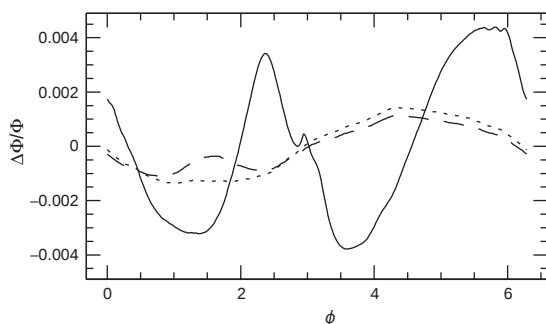


Figure 2 The gravitational potential fluctuations in the disk at angle ϕ at the inner Lindblad resonance where fragmentation first occurs. The amplitude of potential fluctuations in the control run (dotted line) is not significantly increased by the addition of a Jupiter-mass seed planet (dashed line). By $t = 592$ (solid line), the greatly increased planet mass has led to a strong $m = 2$ mode. Fragmentation occurs shortly afterwards. ϕ , gravitational potential.

cases, however, the fragments accrete rapidly from the disk, and so their final masses would fall into the regime of massive planets. Fragmentation also occurs at the same planet mass in a simulation where the initial seed mass is $m_p = 2 \times 10^{-4} m_{\star}$, which is closer to the mass of a giant planet core beginning runaway accretion of disk gas²⁶. For this simulation the time required before fragmentation was roughly doubled. Fragmentation does not occur if we artificially set the mass accreted by a Jupiter mass seed to zero; this verifies that it is the increased planetary mass after significant accretion that leads to instability.

The existence of a propagating mode of planet formation has implications for the evolution of protoplanetary disks and the statistics of planetary systems. In particular, the formation of one massive planet could suffice to trigger rapid planet formation across the range of disk radii for which $Q \approx 1-2$. This could allow massive planets to form at large radii where the timescales for planet formation via other mechanisms can become unfeasibly long compared to typical disk lifetimes. The disruption of the outer disk concomitant with such violent planet formation would allow

the un replenished inner disk to drain viscously onto the star in a short timescale. Studies of the ultraviolet and H α flux arising from the accretion process, and near-infrared flux from the inner disk, suggest that a significant fraction of T Tauri stars are able to dissipate their inner disks rapidly⁷. Related processes may be relevant to the formation of planetary satellite systems¹².

For planetary formation, these results imply that steady growth of giant planets in massive disks around solar mass stars is limited by the vulnerability of the disk to fragmentation once the planetary mass reaches approximately 5 Jupiter masses. The resulting formation of additional planets, which then compete to accrete the available disk gas, implies an upper limit to the mass of massive planets formed by this mechanism. In particular, even if the disk was sufficiently massive it would not be possible to grow a planet from a Jupiter mass far into the brown dwarf regime. This is consistent with observational evidence that planets and brown dwarfs do not share a common mass function¹, which has prompted suggestions that a break in formation mechanisms exists at around 7 Jupiter masses. Finally we note that the endpoint of early disk fragmentation would be a system of numerous massive coplanar planets in initially close-to-circular orbits. Such a system would possess a global organisation imprinted via non-local gravitational effects at birth. These would be favourable initial conditions for the eventual formation of a system comprising one or more massive planets on eccentric orbits⁹. □

Received 25 May; accepted 15 October 1999.

- Mayor, M., Udry, S. & Queloz, D. in *The Tenth Cambridge Workshop on Cool Stars, Stellar Systems and the Sun* (eds Donahue, R. A. & Bookbinder, J. A.) 77–87 (ASP Conf. Ser. 154, San Francisco, 1998).
- Marcy, G. W. & Butler, R. P. Detection of extrasolar giant planets. *Annu. Rev. Astron. Astrophys.* **36**, 57–98 (1998).
- Lin, D. N. C., Bodenheimer, P. & Richardson, D. C. Orbital migration of the planetary companion of 51 Pegasi to its present location. *Nature* **380**, 606–607 (1996).
- Murray, N., Hansen, B., Holman, M. & Tremaine, S. Migrating planets. *Science* **279**, 69–72 (1998).
- Rasio, F. A. & Ford, E. B. Dynamical instabilities and the formation of extrasolar planetary systems. *Science* **274**, 954–965 (1996).
- Weidenschilling, S. J. & Marzari, F. Gravitational scattering as a possible origin for giant planets at small stellar distances. *Nature* **384**, 619–621 (1996).
- Strom, S. E. Initial frequency, lifetime and evolution of YSO disks. *Rev. Mex. Astron. Astrophys. Ser. Conf.* **1**, 317–328 (1995).
- Gladman, B. Dynamics of systems and two close planets. *Icarus* **106**, 247–263 (1993).
- Lin, D. N. C. & Ida, S. On the origin of massive eccentric planets. *Astrophys. J.* **477**, 781–791 (1997).
- Artymowicz, P. & Lubow, S. H. Mass flow through gaps in circumbinary disks. *Astrophys. J.* **467**, L77–L80 (1996).
- Takeuchi, T., Miyama, S. M. & Lin, D. N. C. Gap formation in protoplanetary disks. *Astrophys. J.* **460**, 832–847 (1996).
- Lin, D. N. C. & Papaloizou, J. On the structure of circumbinary accretion discs and the tidal evolution of commensurable satellites. *Mon. Not. R. Astron. Soc.* **188**, 191–201 (1979).
- Kley, W. Mass flow and accretion through gaps in accretion disks. *Mon. Not. R. Astron. Soc.* **303**, 696–710 (1999).
- Bryden, G., Chen, X., Lin, D. N. C., Nelson, R. P. & Papaloizou, J. C. B. Tidally induced gap formation in protostellar disks: Gap clearing and suppression of protoplanetary growth. *Astrophys. J.* **514**, 344–367 (1999).
- Safronov, V. S. *Evolution of the Protoplanetary Cloud and Formation of the Earth and the Planets* (Izdatel'stvo Nauka, Moscow, 1969). (English translation for NASA and NSF by Israel Program for Scientific Translations, Jerusalem, NASA-TT-F-677, 1972).
- Boss, A. P. Evolution of the solar nebula IV. Giant gaseous protoplanet formation. *Astrophys. J.* **503**, 923–937 (1998).
- Osterloh, M. & Beckwith, S. V. W. Millimeter-wave continuum measurements of young stars. *Astrophys. J.* **439**, 288–302 (1995).
- Hayashi, C., Nakazawa, K. & Nakagawa, Y. in *Protostars and planets II* (eds Black, D. C. & Matthews, M. S.) 1100–1153 (Univ. Arizona Press, Tucson, 1985).
- Balbus, S. A., Hawley, J. F. & Stone, J. M. Nonlinear stability, hydrodynamical turbulence, and transport in disks. *Astrophys. J.* **467**, 76–86 (1996).
- Larson, R. B. Gravitational torques and star formation. *Mon. Not. R. Astron. Soc.* **206**, 197–207 (1984).
- Lin, D. N. C. & Pringle, J. E. The formation and initial evolution of protostellar disks. *Astrophys. J.* **358**, 515–524 (1990).
- Toomre, A. On the gravitational stability of a disk of stars. *Astrophys. J.* **139**, 1217–1238 (1964).
- Laughlin, G. & Bodenheimer, P. Nonaxisymmetric evolution in protostellar disks. *Astrophys. J.* **436**, 335–354 (1994).
- Nelson, A. F., Benz, W., Adams, F. C. & Arnett, D. Dynamics of circumstellar disks. *Astrophys. J.* **502**, 342–371 (1998).
- Lissauer, J. J. Timescales for planetary accretion and the structure of the protoplanetary disk. *Icarus* **69**, 249–265 (1987).
- Pollack, J. B. et al. Formation of the giant planets by concurrent accretion of solids and gas. *Icarus* **124**, 62–85 (1996).
- Shakura, N. I. & Sunyaev, R. A. Black holes in binary systems. Observational appearance. *Astron. Astrophys.* **24**, 337–355 (1973).

- Benz, W. in *The Numerical Modelling of Nonlinear Stellar Pulsations* (ed. Buchler, J. R.) 269–287 (Kluwer Academic, Dordrecht, 1990).
- Monaghan, J. J. Smoothed particle hydrodynamics. *Annu. Rev. Astron. Astrophys.* **30**, 543–574 (1992).
- Barnes, J. & Hut, P. A hierarchical O(N log N) force calculation algorithm. *Nature* **324**, 446 (1986).

Acknowledgements

We thank N. Murray for helpful discussions, and N. Wilson for maintaining the required computational resources.

Correspondence and requests for materials should be addressed to P. Armitage (e-mail: armitage@mpa-garching.mpg.de).

The formation of Uranus and Neptune in the Jupiter–Saturn region of the Solar System

Edward W. Thommes*, Martin J. Duncan* & Harold F. Levison†

* Department of Physics, Queen's University, Kingston, Ontario, Canada K7L 3N6

† Space Studies Department, Southwest Research Institute, Boulder, Colorado 80302, USA

Planets are believed to have formed through the accumulation of a large number of small bodies^{1–4}. In the case of the gas-giant planets Jupiter and Saturn, they accreted a significant amount of gas directly from the protosolar nebula after accumulating solid cores of about 5–15 Earth masses^{5,6}. Such models, however, have been unable to produce the smaller ice giants^{7,8} Uranus and Neptune at their present locations, because in that region of the Solar System the small planetary bodies will have been more widely spaced, and less tightly bound gravitationally to the Sun. When applied to the current Jupiter–Saturn zone, a recent theory predicts that, in addition to the solid cores of Jupiter and Saturn, two or three other solid bodies of comparable mass are likely to have formed⁹. Here we report the results of model calculations that demonstrate that such cores will have been gravitationally scattered outwards as Jupiter, and perhaps Saturn, accreted nebular gas. The orbits of these cores then evolve into orbits that resemble those of Uranus and Neptune, as a result of gravitational interactions with the small bodies in the outer disk of the protosolar nebula.

The most plausible model for the formation of giant-planet cores in the Jupiter–Saturn region is based on the concept of ‘oligarchic’ growth⁹. In this model, the largest few objects at any given time are of comparable mass, and are separated by amounts determined by their masses and distances from the Sun. As the system evolves, the mass of the system is concentrated into an ever-decreasing number of bodies of increasing masses and separations. For a given total mass in the system, the model predicts a relationship between the mass of the largest ‘embryos’ and their number.

In the Jupiter–Saturn region, oligarchic growth breaks down when one or more cores begin to accrete a significant amount of nebular gas, thereby significantly increasing their mass(es) in a short period of time. This is expected to occur when the cores reach about 15 Earth masses⁶ (15M $_{\oplus}$), at which time ‘oligarchic’ growth predicts that the cores will be separated by ~ 1 –2 astronomical units (1 AU is the mean Earth–Sun distance). This in turn implies the existence of between 3 and 5 cores in the Jupiter–Saturn zone (4–10 AU).

As the real Solar System has two gas giants and two ice giants, only two of the cores must have accreted a significant amount of gas. Jupiter was possibly the largest core and the first to accrete gas since it is closest to the Sun, where the disk density was highest and the

formation timescales shortest. Perhaps Saturn's core was larger than the others (due to stochastic variations) and also started to accrete gas at this time, or perhaps Saturn did not accrete its gas until later. In either case, the gas-accreting core(s) increased in mass by roughly an order of magnitude in only 10^5 years (ref. 6). At this point oligarchic growth ceased.

We have investigated the subsequent evolution by performing three sets of numerical integrations. In series I, we assume that the Jupiter and Saturn cores accreted their gas simultaneously. We therefore studied the dynamical behaviour of 2 failed cores of $15M_{\oplus}$ (ref. 6), each initially in circular orbits between fully-formed Jupiter and Saturn. In addition to the planetary-size objects, we included a trans-saturnian disk of smaller objects stretching from 10 to 60 AU from the Sun. In series II and III, we investigated the evolution of a system where Jupiter grows first. In series II, we studied the behaviour of four $15M_{\oplus}$ cores distributed between 5 and 9 AU, where we increase the mass of the inner core to that of Jupiter's in 10^5 years. Series III is similar to series II, except that we studied the behaviour of five $15M_{\oplus}$ cores distributed between 6 and 10 AU in order to explore the hypothesis that one core was lost. (See Methods for a complete description of the simulations.)

In all, we performed eight runs in each series. In four of the series I integrations, both failed cores were scattered into the region beyond Saturn and had their orbits circularized due to gravitational interactions with the disk. The temporal evolution of the run that

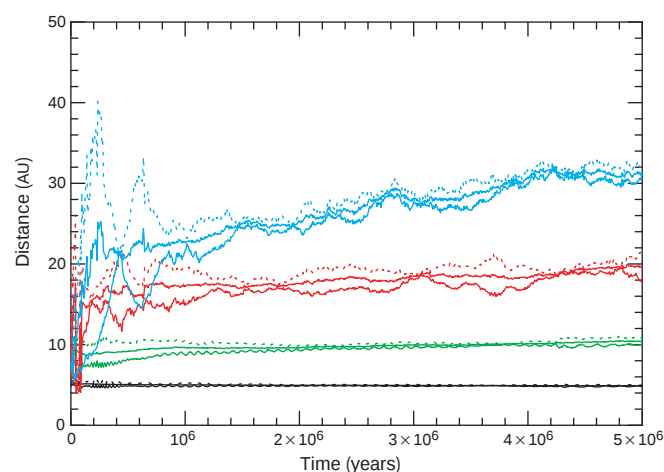


Figure 1 Temporal evolution of model calculations that produced a successful Solar System analogue. Depicted are Jupiter (black), Saturn (green), and the two 'failed cores' (red and blue) in the run which, at its endpoint of 5 Myr, most closely resembles the present Solar System. This run is from series I (see text). Shown are the semi-major axes (thick solid lines) as well as the instantaneous perihelion (thin solid lines) and aphelion (dotted lines) distances of the orbits. Initially the cores suffered multiple gravitational scatterings with Jupiter and Saturn. However, they became decoupled from Jupiter after 9×10^4 years and from Saturn after 2.5×10^5 years. They underwent close encounters with each other until 8×10^5 years. Eventually, the dynamical drag due to small disk objects decoupled them from each other and damped their eccentricities until they were on nearly circular orbits. After this time, they slowly migrated outwards due to the interactions with disk particles. Such a migration has been invoked to explain the structure of the Kuiper belt observed in the real Solar System²¹. At 5×10^6 years the inner core has a semi-major axis, a , of 19.7 AU, an eccentricity, e , of 0.05, and an inclination, i , of 0.2° . For comparison, Uranus currently has $a = 19.2$ AU, $e = 0.006$ and $i = 0.8^\circ$. The outer core has $a = 31.1$ AU, $e = 0.006$ and $i = 0.2^\circ$, and Neptune currently has $a = 30.1$ AU, $e = 0.01$ and $i = 1.7^\circ$. We believe, though, that such a strong correspondence between our model and the real system is largely a coincidence. However, there was one other run in this series that was very similar to the real Solar System (see Fig. S1B in Supplementary Information), so that our mechanism does indeed commonly produce reasonable Solar System analogues. An animation of the dynamical evolution of this system is presented in Supplementary Information.

produced the system that most closely resembles the Solar System is shown in Fig. 1. In seven of the eight series II simulations, all three cores scattered off Jupiter and then evolved onto nearly circular orbits in the outer solar system. In all these cases, one of the cores scattered off Jupiter and evolved onto an orbit near 10 AU, which is the current location of Saturn. That core could subsequently have accreted the amount of nebular gas contained in Saturn. Finally, three of the series III runs produced systems similar to the Solar System. Figure 2 shows 'snapshots' at 5 million years of the system in each series that most closely resembles the Solar System. Similar figures and additional details for all our runs are available; see Supplementary Information.

From our simulations we conclude: first, that this mechanism commonly ($\sim 50\%$) produced reasonable analogues of the real outer planetary system; second, that both 4- and 5-core systems can reasonably reproduce the orbits of the giant planets; and third, that this result seems to be independent of when Saturn accretes its gas.

There is one more issue that must be addressed before we can conclude that we are constructing reasonable analogues of the Solar System. The Solar System is known to have a disk of icy small bodies in nearly circular orbits beyond Neptune, known as the Kuiper belt (shown as small black dots in Fig. 2a; also, see ref. 10 for a review). In order for our mechanism to be applicable to the real Solar System, a similar structure should exist at the end of our simulations. In all our runs there is a population of excited disk objects still encountering the cores (see Fig. 2). These objects should largely be ejected from the Solar System in a few times 10^7 years (ref. 11), and thus are not related to any observed Solar System structure. However, in most of our runs there is a relatively dynamically cold disk that has not been significantly perturbed by the passage of planets—it is this that corresponds to the Kuiper belt. In addition, there is also a population of high-eccentricity objects that are on unstable planet-crossing orbits. This population is similar to the 'scattered disk' recently shown to exist in the Solar System^{12,13}. Thus, our models do indeed produce reasonable Solar System analogues.

A direct comparison between the small-body structures observed in the Solar System and those produced in our simulations is not possible, because there is not yet enough information about the real Kuiper belt. However, we have discovered two previously unknown dynamical processes that helped shape the small-body structures in many of our simulations and which leave well defined dynamical signatures. These may be observed in the real Solar System as the data about the Kuiper belt becomes more complete.

In addition to the scattered disks and 'Kuiper belts' observed in Fig. 2b–d, there is very often a population of excited objects that were perturbed by the planets when the planets were on very eccentric orbits, but which are now beyond the planets' reach. These objects have orbits similar to the scattered disk, but are on stable orbits. As this structure reflects the configuration of the planetary system at early times, we call it the 'fossilized scattered disk'. In addition, in about $\sim 10\%$ of our simulations the objects in the 'Kuiper belt' have had their eccentricities and inclinations excited by the passage of a planet, but not catastrophically so. In these cases, one of the cores was scattered onto a very eccentric, inclined orbit with the semi-major axis $a \geq 60$ AU, which crossed the plane of the system at points both inside and outside the disk, but never actually penetrated the disk. When the core was at the same heliocentric distance as the disk particles it was either above or below the disk due to its inclination. The disk particles were excited by long-range secular effects rather than by gravitational encounters with a planet (see Supplementary Information for some examples).

The idea that the ice giants are failed cores from the Jupiter–Saturn zone represents a significant conceptual shift in our understanding of planet formation. In particular, it suggests that giant-planet formation occurred in a narrow region of the protoplanetary disk, only from ~ 4 AU to ~ 10 AU. Although others have suggested

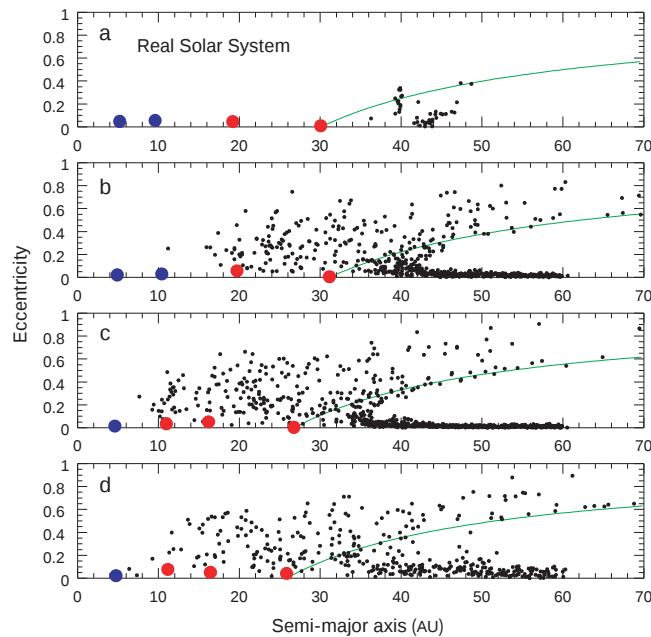


Figure 2 The final orbital elements of most successful runs in each of the three series. For comparison, **a** corresponds to the Solar System, while **b–d** correspond to series I–III, respectively. Eccentricity is plotted versus semi-major axis for the gas giant(s) (blue), the cores (red) and the smaller bodies (black) that originally constitute the planetesimal disk from 10 to 60 AU. These plots were generated 5 million years after the start of the simulation. The green curves show the locus of orbits with perihelion distances at the location of the outer planet. Orbits in the region above and to the left of the green curves are generally unstable on timescales short compared to the age of the Solar System.

that Uranus and Neptune formed somewhat inward of their current locations (for example, Neptune at 23 AU) and gently migrated outwards^{14,15}, the model presented here predicts a significantly more compact region of planet formation. Furthermore, it also suggests that our Solar System went through a stage where Uranus and Neptune were violently removed from their primordial orbits by the giant planets and were thrown out to their current locations. This type of instability has been invoked to explain the large eccentricities seen in some other planetary systems^{16–19}; but we believe that ours is the first model that suggests that a similar upheaval took place in the Solar System. □

Methods

We numerically integrated the orbits of our planets, cores, and trans-planetary planetesimal disk particles for 5 million years using our symplectic integrator, SyMBA²⁰. In series I, we included Jupiter, Saturn and two cores. As we expect Jupiter to migrate slightly inwards and Saturn to move outwards during our integrations^{14,15}, we initially placed them at $a = 5.3$ AU and $a = 9.0$ AU, respectively, with eccentricities and inclinations comparable to their current values. The cores were initially located at $a = 6.35$ AU and $a = 7.6$ AU (as predicted by oligarchic growth) on circular, low-inclination orbits. The runs differed only in our choices of the longitudes of the cores, which were randomly distributed. The series II runs initially had four cores at the same locations. The series III runs initially had 15 $M_{\oplus}$ cores at $a = 5.3, 6.24, 7.36, 8.67$ and 10.21 AU.

To make the simulations numerically tractable, for our trans-planetary disk we used bodies, each of mass $0.24M_{\oplus}$, distributed with a surface density proportional to r^{-1} (series I and II) and $r^{-1.5}$ (series III), where r is heliocentric distance. This corresponds to a total disk mass of 216 $M_{\oplus}$ and 119 $M_{\oplus}$, respectively, a range consistent with estimates of planetary formation efficiency based on our understanding of the mass of the Oort cloud¹⁵. Initially, the disk particles were on nearly circular, low-inclination orbits. Although we included the gravitational effect of the disk particles on the planets and cores, we ignored self-gravity between the disk particles. In addition, we ignored the dynamical-drag effects of gas. Both these effects would tend to assist the disk in circularizing the orbits of the cores. Therefore, we are underestimating the likelihood that the cores will evolve onto circular orbits in the outer solar system.

In our series II and III runs, we replaced Jupiter and Saturn with cores. The mass of Jupiter (and only Jupiter) was linearly increased to its current mass over a period of 10^5 years. In addition, in series II, we extended the protoplanetary disk inward to 4.5 AU by adding a total of 10 $M_{\oplus}$ of material between 4.5 and 10 AU.

Orbits below the green curves are usually stable. An animation of the dynamical evolution of each of these systems is presented in Supplementary Information. In the panel showing the real Solar System, those Kuiper-belt objects which have been observed at multiple oppositions are plotted. The truncation at ~ 48 AU is probably due to observational bias, as more distant objects are less easily detected. We note that 1996 TL₆₆, thus far the only observed member of the scattered disk population observed at multiple oppositions, is not shown as its semi-major axis is 48 AU.

Received 1 July; accepted 6 October 1999.

1. Safronov, V. S. *Evolution of the Protoplanetary Cloud and the Formation of the Earth and Planets* (Nauka, Moscow, 1969).
2. Wetherill, G. W. An alternative model for the formation of the asteroids. *Icarus* **100**, 307–325 (1992).
3. Chambers, J. E. & Wetherill, G. W. Making the terrestrial planets: N-body integrations of planetary embryos in three dimensions. *Icarus* **136**, 304–327 (1999).
4. Agnor, C. B., Canup, R. & Levison, H. On the character and consequences of large impacts in the late stage of terrestrial planet formation. *Icarus* **142**, 219–237 (1999).
5. Bodenheimer, P. & Pollack, J. B. Calculations of the accretion and evolution of giant planets: the effects of solid cores. *Icarus* **67**, 391–408 (1986).
6. Pollack, J. B. et al. Formation of the giant planets by concurrent accretion of solids and gas. *Icarus* **124**, 62–85 (1996).
7. Lissauer, J. J., Pollack, J. B., Wetherill, G. W. & Stevenson, D. J. in *Neptune and Triton* (ed. Cruikshank, D. P.) 37–108 (Univ. Arizona Press, Tucson, 1996).
8. Stewart, G. & Levison, H. F. On the formation of Uranus and Neptune. *Proc. Lunar Planet. Sci. Conf. XXIX*, abstr. no. 1960 (1998).
9. Kokubo, E. & Ida, S. Oligarchic growth of protoplanets. *Icarus* **131**, 171–178 (1998).
10. Morbidelli, A. in *Solar System Formation and Evolution* (eds. Lazzaro, D., Vieira Martins, R., Ferraz-Mello, S., Fernandez, J. & Beauge, C.) 83 (ASP Conf. Ser. 149, 1998).
11. Levison, H. & Duncan, M. From the Kuiper belt to Jupiter-family comets: the spatial distribution of ecliptic comets. *Icarus* **127**, 13–32 (1997).
12. Duncan, M. & Levison, H. F. A scattered disk of icy objects and the origin of Jupiter-family comets. *Science* **276**, 1670–1672 (1997).
13. Luu, J. et al. A new dynamical class of object in the outer Solar System. *Nature* **387**, 573–575 (1997).
14. Fernandez, J. A. & Ip, W.-H. Some dynamical aspects of the accretion of Uranus and Neptune—The exchange of orbital angular momentum with planetesimals. *Icarus* **58**, 109–120 (1984).
15. Hahn, J. & Malhotra, R. Radial migration of planets embedded in a massive planetesimal disk. *Astron. J.* **117**, 3041–3053 (1999).
16. Rasio, F. A. & Ford, E. B. Dynamical instabilities and the formation of extrasolar planetary systems. *Science* **274**, 954–956 (1996).
17. Weidenschilling, S. J. & Marzari, F. Gravitational scattering as a possible origin for giant planets at small stellar distances. *Nature* **384**, 619–621 (1996).
18. Lin, D. N. C. & Ida, S. On the origin of massive eccentric planets. *Astrophys. J.* **477**, 781–791 (1997).
19. Levison, H. F., Lissauer, J. J. & Duncan, M. J. Modeling the diversity of outer planetary systems. *Astron. J.* **116**, 1998–2014 (1998).
20. Duncan, M., Levison, H. & Lee, M. H. A multiple timestep symplectic algorithm for integrating close encounters. *Astron. J.* **116**, 2067–2077 (1998).
21. Malhotra, R. The origin of Pluto's orbit: implications for the Solar System beyond Neptune. *Astron. J.* **110**, 420–429 (1995).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank G. Stewart and K. Zahnle for discussions, and L. Dones, B. Gladman, W. McKinnon, J. Parker, A. Stern, W. Ward and G. Wetherill for comments on an earlier draft. E.W.T. and M.J.D. were supported by the Natural Science and Engineering Research Council of Canada, H.E.L. was supported by NASA's Planetary Geology & Geophysics, Origins of Solar Systems, and Exobiology programme.

Correspondence and requests for materials should be addressed to M.D. (e-mail: duncan@astro.queensu.ca).

First-order phase transitions in a quantum Hall ferromagnet

Vincenzo Piazza*, Vittorio Pellegrini*, Fabio Beltram*, Werner Wegscheider†, Tomáš Jungwirth‡§ & Allan H. MacDonald‡

* Scuola Normale Superiore and Istituto Nazionale per la Fisica della Materia, I-56126 Pisa, Italy

† Walter Schottky Institute, Munich, Germany

‡ Department of Physics, Indiana University, Bloomington, Indiana 47405, USA

§ Institute of Physics ASCR, Cukrovarnická 10, 162 00 Praha 6, Czech Republic

The single-particle energy spectrum of a two-dimensional electron gas in a perpendicular magnetic field consists of equally spaced energy states, known as Landau levels. Each level is split owing to spin interactions, and its degeneracy is proportional to the magnetic field strength. When the ratio, ν (or 'filling factor'), of the number of electrons and the degeneracy of a Landau level takes an integer or particular fractional values, quantum Hall effects¹ occur, characterized by a vanishingly small longitudinal resistance and a quantized (transverse) Hall voltage². The quantum Hall regime may be used for the controlled study of many-particle cooperative phenomena, such as order-disorder phase transitions (analogous to those observed in conventional magnets). Both isotropic and anisotropic ferromagnetic ground states have been predicted^{3–8} to occur in the quantum Hall regime, some of which have been investigated experimentally^{9–13} in samples with different geometries and filling factors. Here we report evidence for first-order phase transitions in quantum Hall states ($\nu = 2, 4$) confined to a wide gallium arsenide quantum well. We observe hysteresis and an anomalous temperature dependence in the longitudinal resistivity, indicative of a transition between two distinct ground states of an Ising quantum Hall ferromagnet. The microscopic origin of the anisotropy field is identified using detailed many-body calculations.

The study of quantum phase transitions^{14,15}, 'tuned' by system parameters rather than temperature, has enlarged the domain of phase-transition physics. In electronic systems, transitions to ordered ground states can occur when interactions between particles—rather than single-particle potentials—play a dominant role in selecting the ground state. This situation is frequently encountered in the quantum Hall regime owing to the quantization of in-plane kinetic energy into macroscopically degenerate Landau levels. The ordered many-particle ground states discussed originate when two or, in general, N Landau levels are brought close to alignment and the total filling factor of these levels is an integer less than $N^{3,6–8}$. The close analogy between these states and those of two-dimensional ferromagnets is often emphasized by using a pseudospin language¹⁶ to describe the Landau level degree of freedom. For the two-level case, we refer to one of the single-particle Landau levels as the pseudospin-up state and to the other as the pseudospin-down state. A general quantum spinor with an arbitrary orientation can be obtained as a linear combination of the up and down states.

Ferromagnetic states are characterized by non-zero pseudospin polarization of the two-component many-particle system, even when the splitting between the two Landau levels vanishes. In single-layer two-dimensional systems at $\nu = 1$, the pseudospin degree of freedom can be the real electron spin. In this case it is rigorously established³ that the ground state is an isotropic strong ferromagnet. In general, the pseudospin degree of freedom can involve real-spin, the cyclotron orbit-radius quantum number, and the layer index in multiple quantum wells or the sub-band index in wide quantum wells⁵. Electron-electron interactions are then expected to lead to tunable pseudospin anisotropy and to a complex range of ordered states which can be explored by adjusting system parameters, often in the same physical sample. In a double-quantum-well structure, for example, Landau level crossings can be accompanied by the introduction of a broken-symmetry phase with easy-plane pseudospin anisotropy whose soft collective excitations and related quantum phase transitions were recently observed^{11–13}.

The possibility of realizing quantum Hall ferromagnets with the easy-axis (spin aligning along some fixed direction below the critical temperature)—Ising—anisotropy, common in familiar magnetic systems, has also been raised⁵. Experiments¹⁷ on several fractional quantum Hall effect (QHE) transitions suggested interpretations within this framework. However, no rigorous physical picture has been developed to date which is able to describe quantum Hall ferromagnets at fractional filling factors. The experiments reported here reveal intriguing phenomena arising when Landau levels with opposite spin and different sub-band indices are brought close to degeneracy by applying an external electric field. The observed

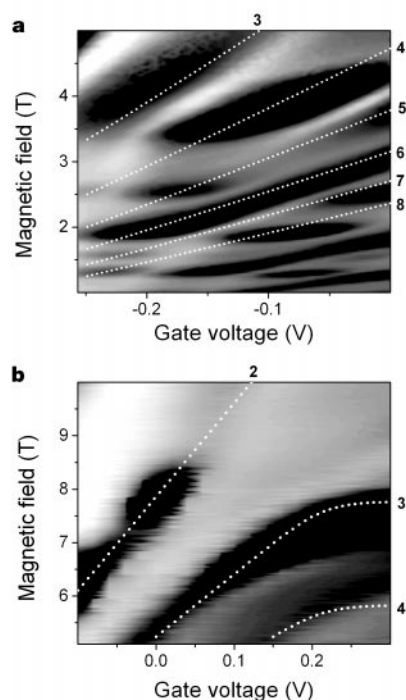


Figure 1 Longitudinal resistivity as a function of the magnetic field and gate voltage. Dark (bright) regions represent low (high) resistivity values. The dotted lines show the evolution of the quantum-Hall-effect minima and are labelled by the corresponding filling factors ν . For gate voltages $V_g > -0.26$ V, two sub-bands are occupied: the complex resistivity pattern in this region originates from crossings between Landau levels of the two sub-bands. Similar evolution patterns of the resistivity are obtained in different measurement sessions after thermally cycling the sample. Changes of gate voltages of ~ 20 mV are observed in different measurement sessions owing to small differences in the total electron density. Measurements shown in both **a** and **b** were performed at $T = 330$ mK, with lock-in techniques at a frequency of 12.5 Hz; the a.c. excitation current was set to 100 nA to avoid electron heating.

suppressions of the $\nu = 2$ and 4 QHEs and hysteretic behaviour in the low-temperature longitudinal resistance indicate a first-order phase transition between oppositely polarized ground states of an Ising ferromagnet with domain structure. Detailed many-body calculations confirm that at external fields consistent with experiment the ground state of the system is a quantum Hall ferromagnet with easy-axis anisotropy.

The sample grown for this study consists of a 60-nm-wide gallium arsenide (GaAs) quantum well positioned 75 nm below the sample surface and embedded between two hard-wall $\text{Al}_{0.25}\text{Ga}_{0.75}\text{As}$ barriers. An important feature of our system is a soft barrier originating from Coulomb interactions among electrons in the quantum well. This barrier separates the electronic system into two well defined, yet strongly coupled, layers. We argue below that the soft barrier is the source of the anisotropy field leading to Ising ferromagnetism in our sample. The two lowest energy sub-bands of the unbiased quantum well are occupied by electrons. When a perpendicular magnetic field is applied, these sub-bands form two distinct Landau-level ladders, each consisting of alternating spin-up ($\uparrow$) and spin-down ($\downarrow$) Landau levels labelled by the orbit-radius quantum number n . A gate contact evaporated on the surface of the sample allows us to apply an electric field across the quantum well. This external electric field changes the two-dimensional electronic density in the well and, more importantly, varies the separation between the two lowest energy levels in the well. At bias potential $V_g = 0$, the electron density is $5.5 \times 10^{11} \text{ cm}^{-2}$ and the mobility $2.6 \times 10^6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. In our experiments, we use the external bias potential to align different pairs of Landau levels.

In Fig. 1a and b we show the measured longitudinal resistivity ρ_{xx} as a function of magnetic field B and applied bias, at temperature $T = 330 \text{ mK}$; dark (bright) regions correspond to low (high) resistivity. At a given gate voltage, the minima of ρ_{xx} as a function of B correspond to integer filling factors. The evolution of the minima can be traced in the figure by following the dotted lines labelled by the corresponding Landau level filling factor. For the gate bias values of Fig. 1 ($V_g > -0.26 \text{ V}$), both Landau-level ladders are populated and ρ_{xx} exhibits a complicated pattern reflecting the crossing of Landau levels from different ladders.

In order to study phase transitions in our system, we now focus on the even-integer filling factors $\nu = 2$ and $\nu = 4$. The evolution of the $\nu = 4$ state as the gate voltage is swept from $V_g = -0.06 \text{ V}$ to $V_g = -0.19 \text{ V}$ is shown in Fig. 2d. The quantum Hall minimum in ρ_{xx} disappears near the predicted critical biases at which a filled Landau level crosses with an empty level (see Fig. 2 legend). We note that the suppression of the quantum Hall minimum in the longitudinal resistivity correlates with the emergence of hysteresis in ρ_{xx} in up and down sweeps of the magnetic field. Representative curves are shown in Fig. 3a for $T = 330 \text{ mK}$. This figure indicates that the detailed shape and position of the observed hysteresis is sensitive to small changes in the electron density which follow each thermal cycling of the sample; differences in the observed critical bias potentials at which the QHE is suppressed and hysteresis emerges correspond to the density variation in different measurement sessions.

Figure 3b shows an enlarged view of the peak near the $\nu = 4$ minimum of ρ_{xx} ; here the hysteresis extends over a large region of magnetic fields between $\nu = 4$ and $\nu = 3$ states and occurs at $V_g = -0.152 \text{ V}$. The temperature evolution of the hysteresis (Fig. 3c) confirms that the characteristic energy scale involved in this phenomenon is that of the electron-electron interaction, as the hysteresis disappears at temperatures T_c above the estimated⁵ Ising critical temperature for a quantum Hall ferromagnet, $\sim 1 \text{ K}$. Thermal cycling of the sample was found to have a negligible effect on measured T_c .

A similar phenomenology applied to the $\nu = 2$ quantum Hall state evolution (Fig. 2c). As for the $\nu = 4$ case, the onset of hysteresis

(see Fig. 2 inset) and the suppression of the quantum Hall state are strictly correlated. The transport anomalies described above have not been observed at odd-integer filling factors. We attribute the appearance of hysteresis to the development of pseudospin magnetic order in the two-dimensional Ising universality class, and explain below why, in our sample, easy-axis pseudospin anisotropy is expected for even-integer filling factors.

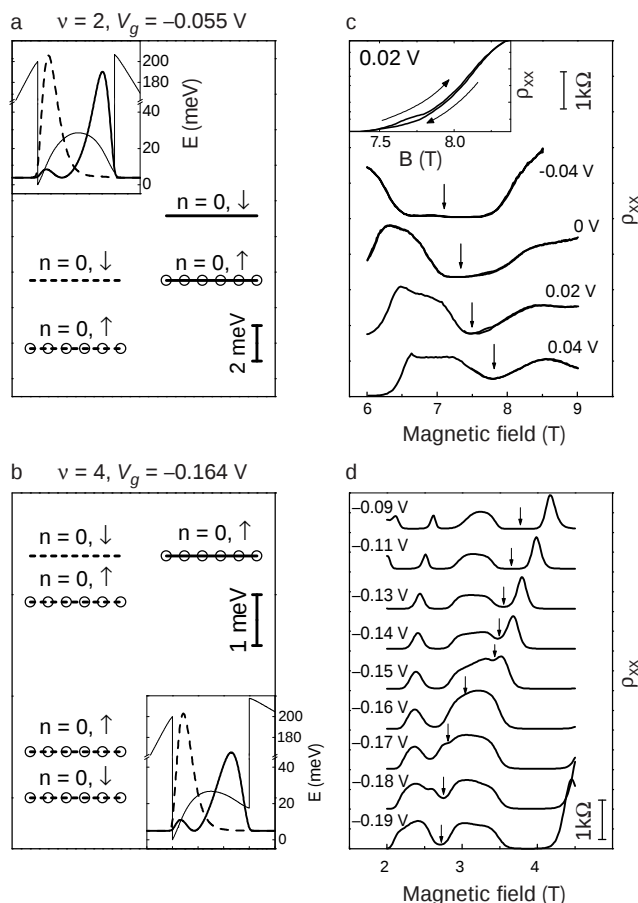


Figure 2 Energy spectra and resistivity behaviour for $\nu = 2, 4$. **a, b**, Self-consistent local-density-approximation (LDA) calculations of the energy spectrum for $\nu = 2$ (**a**) and $\nu = 4$ (**b**). Circles mark the occupied states, insets show LDA quantum-well profiles and wavefunctions. Landau levels corresponding to states that occupy dominantly the right (left) hand side of the well are labelled by index $R(L)$. For $\nu = 2$, the filled $n = 0, \uparrow, R$ level (n is the orbit-radius quantum number and $\uparrow, \downarrow$ label spin up, spin down) is brought to degeneracy with the empty $n = 0, \downarrow, L$ level at a critical potential $V_{\nu=2}^{(c)} = -0.055 \text{ V}$. The second critical potential $V_{\nu=2}^{(c2)} = 0.01 \text{ V}$ corresponds to the alignment (not shown in figure) of filled $n = 0, \uparrow, L$ and empty $n = 0, \downarrow, R$ Landau levels. At $\nu = 4$, the critical bias potentials are $V_{\nu=4}^{(c1)} = -0.164 \text{ V}$ and $V_{\nu=4}^{(c2)} = -0.144 \text{ V}$. We note that the LDA Landau-level energies depend on which Landau levels are assumed to be occupied. For example, at $\nu = 2$, the spin-polarized configuration with both R and L Landau ladders occupied, and the spin-unpolarized case with electrons only in one of the ladders, are both self-consistent LDA solutions in the interval between $V_{\nu=2}^{(c1)}$ and $V_{\nu=2}^{(c2)}$. This is consistent with the observed hysteretic behaviour. At $\nu = 4$, on the other hand, no self-consistent LDA ground state can be found when the bias is swept from approximately -0.1 V to $V_{\nu=4}^{(c2)}$ or from $V_{\nu=4}^{(c1)}$ to approximately -0.2 V . Both bistability and failure to achieve self-consistency in the mean-field-like LDA calculations signal the importance of correlations in determining the many-particle ground state. **c, d**, Set of curves of resistivity (ρ_{xx}) versus magnetic field, recorded at different gate voltages and $T = 330 \text{ mK}$, showing the disappearance of the $\nu = 2$ (**c**) and $\nu = 4$ (**d**) quantum-Hall states (marked with arrows). For the $\nu = 2$ case, the resistivity for up and down sweeps of the magnetic field is also shown. In the inset an enlarged view of the trace for $V_g = 0.02 \text{ V}$ is shown. Hysteretic behaviour appears when the gap responsible for the quantum Hall state is strongly reduced. The curves are offset for clarity, and each curve is labelled with the gate voltage.

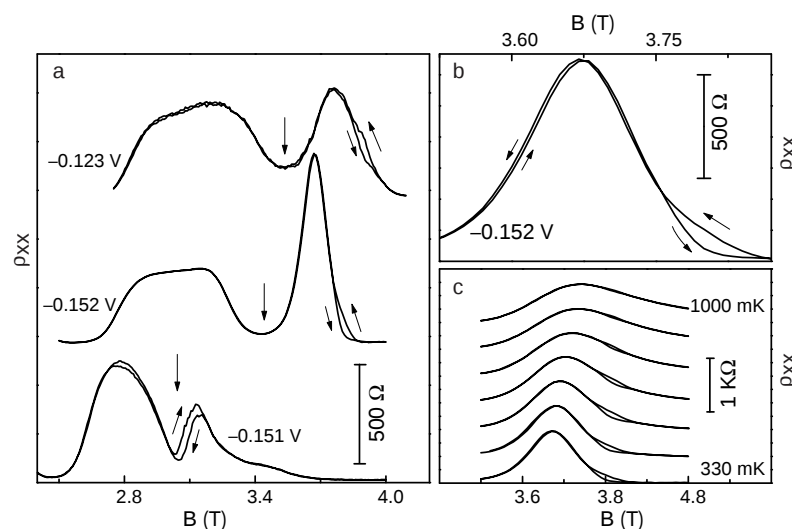


Figure 3 Resistivity traces. **a**, Traces of longitudinal resistivity ρ_{xx} at $T = 330$ mK near filling factor $\nu = 4$, obtained after thermally cycling the sample, for up and down sweeps of the magnetic field (B). The up and down sweeps are indicated by arrows approximately parallel to the trace. The presence of hysteresis is always correlated to the suppression of the $\nu = 4$ state (indicated by vertical arrows). The shift in the position of

the $\nu = 4$ state indicates small differences (less than 15%) in the total electron density obtained after thermally cycling the sample. **b**, Enlarged view of the trace shown in **a** at $V_g = -0.152$ V; the observed hysteresis extends over a large range of magnetic fields between $\nu = 4$ and $\nu = 3$. **c**, Temperature evolution of the hysteresis at $V_g = -0.152$ V. The curves are offset for clarity.

For a disorder-free system at $T = 0$, easy-axis pseudospin anisotropy implies that external fields exceeding a coercivity field have to be applied in the direction opposite to the pseudospin orientation in order to produce the pseudospin reversal. This property leads to hysteretic behaviour of the pseudospin orientation and first-order phase transitions at the extrema of the hysteresis loop. In real systems it is expected that disorder will produce a random effective external field, leading to the formation of domains with particular pseudospin orientations. Dissipation due to mobile charges created at domain walls, where the quasiparticle excitation gap is reduced, can lead to a breakdown of the quantum Hall effect as shown in Figs 1 and 2. In the disorder-free limit, long-range order is destroyed at a critical temperature yielding a finite temperature continuous phase transition in the Ising universality class.

The hysteresis and anomalous temperature dependence of the $\nu = 2$ and $\nu = 4$ ρ_{xx} traces presented above fall into the class of effects expected for easy-axis ferromagnets with a domain structure. The knowledge of the physical meaning of the pseudospin degree of freedom allows us to identify the microscopic origin of the easy-axis anisotropy in our sample. Assuming a many-body fermionic wavefunction constructed from one-particle states with common pseudospin orientation θ , we calculated the expectation value of the many-body hamiltonian. ($\theta = 0$ corresponds to pseudospin up, and $\theta = \pi$ to pseudospin down.) We focus on the simpler case of $\nu = 2$. We consider first a model system, consisting of two infinitely narrow quantum wells separated by a rigid tunnelling barrier. Within this model, the pseudospin dependent contribution to the ground state energy at small V_g is given by^{5,8} $E_\theta = (\Delta_z - \Delta_T - V_g^2/\Delta_T)\cos\theta$, and the critical gate potentials by $V^{(1)} = -V^{(2)} = -\sqrt{\Delta_z\Delta_T - \Delta_T^2}$. In this expression, Δ_T is the tunnelling gap, that is, the energy splitting between the two lowest quantum levels in the unbiased double-well, and Δ_z is the Zeeman splitting. At this level of approximation, the pseudospin ferromagnet is isotropic since there is no preferred pseudospin orientation when $V_g = V^c$: $E_\theta(V_g = V^c) \equiv 0$. To give rise to easy-axis anisotropy, we must take into account the softness of the barrier separating the two layers. At, for example, negative bias, the $n = 0, \uparrow, L$ Landau level (R and L label the ladder whose wavefunction has dominant weight near right- and left-hand side of the quantum well, respectively) is always fully occupied, and so does not contribute to the anisotropy energy. Here $n = 0, \uparrow, L$ and

$n = 0, \uparrow, R$ are the two pseudospin levels. We find in our local-density-approximation (LDA) calculations that at sufficiently high electron densities—consistent with the experimental density at which the $\nu = 2$ transition occurs—the tunnelling barrier is reduced when the $n = 0, \downarrow, L$ level is occupied, compared to the case of the occupied $n = 0, \uparrow, R$ Landau level. In the LDA calculations this occurs because the exchange energy (which counters the electrostatic barrier between layers) is larger in the latter, fully spin-polarized, case.

From a many-body point of view, the reduction in the quasiparticle tunnelling gap reflects the correlated mixing of higher electronic sub-bands into many-body states. Smaller barriers lead to a larger tunnelling gap, that is, to a larger separation between L and R Landau levels. Translated into the pseudospin language, the softness of the barrier leads to the increase of the gap between the occupied lower-energy pseudospin level and the empty higher-energy pseudospin level in either up or down pseudospin polarizations. We can account for this property by replacing Δ_T in the ground state energy expression given above, by $\Delta_T^0 + \delta\Delta_T \cos\theta$, where $\delta\Delta_T > 0$. The anisotropy energy term $-\delta\Delta_T \cos^2\theta$ favours the pseudospin angle $\theta = 0$ or π , that is, the system behaves like an Ising ferromagnet. Similar arguments explain easy-axis anisotropy at filling factor $\nu = 4$. For odd-integer filling factors, the two pseudospin Landau levels have the same real-spin and the above scheme does not apply. The odd-integer quantum Hall effect states have easy-plane anisotropy consistent with our experimental observation.

These unusual magnetotransport observations are but one instance of a broad class of phenomena that can be associated with crossing Landau levels. Even for the relatively simple case of two crossing levels, broken-symmetry ground states analogous to those of isotropic, easy-plane or easy-axis two-dimensional ferromagnets can occur. The phenomenology of easy-axis systems, dominated by the hysteretic effects and domain-morphology issues familiar in common magnetic materials, is particularly rich. The observations presented here suggest that quantum Hall systems can be ideal tools for the study of these phenomena. □

Received 4 June; accepted 20 October 1999.

1. Prange, R. E. & Girvin, S. M. (eds) *The Quantum Hall Effect*, 2nd edn (Springer, New York, 1990).
2. Das Sarma, S. & Pinczuk, A. (eds) *Perspectives in Quantum Hall Effects* (Wiley, New York, 1996).

3. Girvin, S. M. & MacDonald, A. H. in *Perspectives in Quantum Hall Effects* (eds Das Sarma, S. & Pinczuk, A.) 161–224 (Wiley, New York, 1996).
4. Giuliani, G. F. & Quinn, J. J. Spin-polarization instability in a tilted magnetic field of a two-dimensional electron gas with filled Landau levels. *Phys. Rev. B* **31**, 6228–6232 (1985).
5. Jungwirth, T., Shukla, S. P., Smrčka, L., Shayegan, M. & MacDonald, A. H. Magnetic anisotropy in quantum Hall ferromagnets. *Phys. Rev. Lett.* **81**, 2328–2331 (1998).
6. Zheng, L., Radtke, R. J. & Das Sarma, S. Spin-excitation-instability-induced quantum phase transitions in double-layer quantum Hall systems. *Phys. Rev. Lett.* **78**, 2453–2456 (1997).
7. Das Sarma, S., Sachdev, S. & Zheng, L. Double-layer quantum Hall antiferromagnetism at filling fraction $\nu = 2/m$ where m is an odd integer. *Phys. Rev. Lett.* **79**, 917–920 (1997).
8. MacDonald, A. H., Rajaraman, R. & Jungwirth, T. Broken symmetry ground states in $\nu = 2$ bilayer quantum Hall systems. *Phys. Rev. B* **60**, 8817–8826 (1999).
9. Eisenstein, J. P. *Perspectives in Quantum Hall Effects* (eds Das Sarma, S. & Pinczuk, A.) 58–70 (Wiley, New York, 1996).
10. Daneshvar, A. J. *et al.* Magnetisation instability in a two-dimensional system. *Phys. Rev. Lett.* **79**, 4449–4452 (1997).
11. Pellegrini, V. *et al.* Collapse of spin excitations in quantum Hall states of coupled electron double layers. *Phys. Rev. Lett.* **78**, 310–313 (1997).
12. Pellegrini, V. *et al.* Evidence of soft-mode quantum phase transitions in electron double layers. *Science* **281**, 799–802 (1998).
13. Sawada, A. *et al.* Phase transition in the $\nu = 2$ bilayer quantum Hall state. *Phys. Rev. Lett.* **80**, 4534–4537 (1998).
14. Sachdev, S. *Quantum Phase Transitions* (Cambridge Univ. Press, 1999).
15. Sondhi, S. L., Girvin, S. M., Carini, J. P. & Shahar, D. Continuous quantum phase transitions. *Rev. Mod. Phys.* **69**, 315–333 (1997).
16. MacDonald, A. H., Platzman, P. M. & Boebinger, G. S. Collapse of integer Hall gaps in a double-quantum-well system. *Phys. Rev. Lett.* **65**, 775–778 (1990).
17. Cho, H. *et al.* Hysteresis and spin transitions in the fractional quantum Hall effect. *Phys. Rev. Lett.* **81**, 2522–2525 (1998).

Acknowledgements

We thank M. Bichler for technical support during sample growth, and G. F. Giuliani for discussions. The work at Scuola Normale Superiore was funded in part by MURST. The work at Indiana University was supported by the NSF, and at Institute of Physics ASCR by the Ministry of Education of the Czech Republic and the Grant Agency of the Czech Republic.

Correspondence and requests for materials should be addressed to V. Piazza (e-mail: piazza@cmp.sns.it).

Phase-coherent amplification of atomic matter waves

S. Inouye, T. Pfau, S. Gupta, A. P. Chikkatur, A. Görlitz, D. E. Pritchard & W. Ketterle

Department of Physics and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Atomic matter waves, like electromagnetic waves, can be focused, reflected, guided and split by currently available passive atom-optical elements. However, the key for many applications of electromagnetic waves lies in the availability of amplifiers. These active devices allow small signals to be detected, and led to the development of masers and lasers. Although coherent atomic beams have been produced^{1–4}, matter wave amplification has not been directly observed. Here we report the observation of phase-coherent amplification of atomic matter waves. The active medium is a Bose–Einstein condensate, pumped by light that is far off resonance. An atomic wave packet is split off the condensate by diffraction from an optical standing wave, and then amplified. We verified the phase coherence of the amplifier by observing interference of the output wave with a reference wave packet. This development provides a new tool for atom optics and atom interferometry, and opens the way to the construction of active matter-wave devices.

Matter-wave amplification differs from light amplification in one important respect. Because the total number of atoms is conserved (in contrast to photons), the active medium of a matter-wave amplifier has to include a reservoir of atoms. Also needed is a

coupling mechanism that transfers atoms from the reservoir to the input mode while conserving energy and momentum. Amplification is realized if the transfer mechanism is accelerated by the build-up of atoms in the final state and inversion is maintained between the initial and final states.

It is convenient to use a Bose–Einstein condensate with its high phase-space density to realize matter-wave amplification, although in principle it is not necessary^{1–4}. This is analogous to the situation in optics where coherent light is used for pumping parametric amplifiers. The momentum required to transfer atoms from the condensate at rest to the input mode can be provided by light scattering. As was discussed in refs 5 and 6, a condensate pumped by an off-resonant laser acts as a matter-wave amplifier. It can amplify input matter waves within the momentum range that can be reached by scattering a single pump photon. Energy is conserved because the scattered light has lower frequency.

The inversion in this matter-wave amplifier is most apparent in the ‘dressed atom’ picture, where the condensate at rest and the pump light field form an upper state that can decay into recoiling atoms and scattered photons. The escape of the photons maintains inversion, allowing, in principle, a complete transfer of the condensate atoms into the recoil mode.

The gain process can be explained in a semiclassical picture. The input matter wave of wavevector $\mathbf{K}_j$ interferes with the condensate at rest, and forms a moving matter-wave grating; this grating diffracts the pump light with wavevector $\mathbf{k}_0$ into the momentum and energy conserving direction $\mathbf{k}_i = (\mathbf{k}_0 - \mathbf{K}_j)$. The momentum imparted by the photon scattering is absorbed by the matter-wave grating by coherently transferring an atom from the condensate into the recoil mode, which is the input mode. The diffraction efficiency of the grating is proportional to the square of the depth of the density modulation, and therefore to the number of atoms in the input mode, N_j . This implies an exponential growth of N_j (as long as one can neglect the depletion of the condensate at rest).

The amplification of atoms in a recoil mode j follows a gain equation^{7,8}

$$\dot{N}_j = (G_j - L_j)N_j \quad (1)$$

with the gain coefficient

$$G_j = RN_0 \frac{\sin^2 \theta_j}{8\pi/3} \Omega_j \quad (2)$$

Here R is the rate for single-atom Rayleigh scattering which is proportional to the pump light intensity, N_0 is the number of atoms in the condensate at rest, θ_j is the angle between the polarization of the incident light and the direction of the scattered light, and Ω_j is the phase-matching solid angle for scattering into mode j . The loss term L_j describes the decoherence rate of the matter-wave grating and determines the threshold for exponential growth (see ref. 7 for details). We have recently observed the build-up of superradiant emission due to this gain, and have confirmed the salient features of the model presented above. Previous theoretical discussions of this system had assumed an optical cavity into which the pump light is scattered^{5,6}, but our experiment⁷ showed that cigar-shaped condensates have sufficient gain for axial light emission, even in free space. In our observation of superradiance, there were no input atoms, and the build-up of a macroscopic matter wave was initiated by spontaneous scattering. Here we characterize the amplification process by providing a variable input and measuring the amplitude and phase of the amplified matter wave.

Cigar-shaped Bose–Einstein condensates of a few million sodium atoms in the $F = 1$, $m_F = -1$ state were produced in a magnetic trap using the standard techniques of laser cooling and evaporative cooling⁹. The condensate was 200 μm in length and 18 μm in diameter. Input matter waves with a well defined momentum were generated by exposing the condensate to a pulsed optical standing wave which transferred a small fraction of the atoms (10^{-4} to 10^{-2}) into a recoil mode by Bragg diffraction^{10,11}. The

frequencies of both laser beams were red-detuned by 1.7 GHz from the $3S_{1/2}, F = 1 \rightarrow 3P_{3/2}, F = 0, 1, 2$ transition, and the intensities were kept below the threshold for superradiant Rayleigh scattering⁷. The geometry of the light beams is shown in Fig. 1. The beam which was perpendicular to the long axis of the condensate (the radial beam) was blue-detuned by 50 kHz relative to the axial beam. This detuning fulfills the Bragg resonance condition; that is, it corresponds to the kinetic energy of the recoiling atoms. The broadening of the Bragg resonance by the finite pulse length (10 μ s) was small enough to prevent the off-resonant excitation of other diffraction orders.

Amplification of the input matter wave was realized by applying an intense pump pulse along the direction of the radial Bragg beam for the next 20 μ s with a typical intensity of 40 mW cm⁻². The number of atoms in the recoil mode was determined by suddenly switching off the trap and observing the ballistically expanding cloud of atoms after 35 ms using resonant absorption imaging. After the expansion, the condensate and the recoiling atoms were fully separated (Fig. 2a–c).

Figure 2 shows the input–output characteristics of the amplifier. The number of input atoms was below the detection limit of our absorption imaging (Fig. 2a) and was determined from a calibration of the Bragg process at high laser powers, where the diffracted atoms were clearly visible in the images. For short pulses, the diffraction efficiency was quadratic in the laser power, as expected. The amplification pulse alone, although above the threshold for superradiance⁷, did not generate a discernible signal of atoms in the recoil mode (Fig. 2b). When the weak input matter wave was added, the amplified signal was clearly visible (Fig. 2c). The gain was controlled by the intensity of the pump pulse (see equation (2)) and typically varied between 10 and 100. Figure 2d shows the observed linear relationship between the atom numbers in the input and the amplified output with a number gain of 30.

The phase of the amplified matter wave was determined with an interferometric technique. For this, a reference matter wave was split off the condensate in the trap in the same manner as the first (input) wave. The phase of the reference matter wave was scanned by shifting the phase of the radio-frequency signal that drove the acousto-optic modulator generating the axial Bragg beam. We then observed the interference between the reference and the amplified matter waves by measuring the number of atoms in the recoil mode.

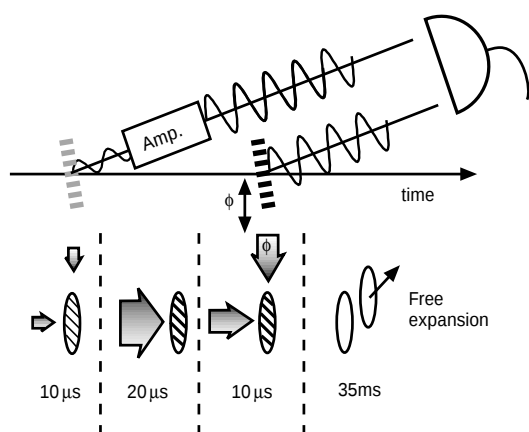


Figure 1 Experimental scheme for observing phase-coherent matter-wave amplification. A small-amplitude matter wave was split off the condensate by applying a pulse of two off-resonant laser beams (Bragg pulse). This input matter wave was amplified by passing it through the condensate pumped by a laser beam. The coherence of the amplified wave was verified by observing its interference with a reference matter wave, which was produced by applying a second (reference) Bragg pulse to the condensate. The interference signal was observed after 35 ms of ballistic expansion. Note that the polarization of the radial beam was perpendicular to the long axis of the condensate ($\theta = \pi/2$). Extinction of the axial beam during the amplifying pulse was better than 10^{-8} .

If we characterize the two interfering matter waves by their atom numbers N_i and normalized wavefunctions ψ_i , then the visibility V of the interference pattern is given by:

$$V = \frac{2\langle\psi_1|\psi_2\rangle\sqrt{N_1N_2}}{N_1 + N_2} \quad (3)$$

$$\sim 2\langle\psi_1|\psi_2\rangle\sqrt{N_1/N_2} \quad (N_1 \ll N_2) \quad (4)$$

This simple description is valid for small transfer efficiencies of the Bragg pulses ($N_i \ll N_0$). A more general description treats the two Bragg pulses as a Ramsey-type matter-wave interferometer¹², where the phase of the matter-wave grating (formed by the recoiling atoms and the condensate) is compared with the phase of the standing-wave light grating used for the second Bragg transition. This treatment reproduces equation (3), where N_1 is replaced by $N_1(1 - N_2/N_0)$ and N_2 by $N_2(1 - N_1/N_0)$.

In Fig. 3 we show the interference signal between the amplified input and the reference matter wave as the reference phase was scanned. When the amplitude of the input was comparable to that of the reference matter wave, high-contrast fringes were observed even without amplification (Fig. 3a). When the input was about 40 times weaker in population, fringes were barely visible (Fig. 3b). After amplification, we regained a large visibility (Fig. 3c). This increase in visibility proves the coherent nature of the matter-wave amplification process.

The visibility was studied as a function of the number of input atoms (Fig. 4). Since amplification did not produce any observable phase shift, the visibility could be measured by comparing the number of atoms in the recoil mode at constructive and destructive phases. Without amplification, the visibility increased with the number of input atoms in fair agreement with equation (3), assuming $\langle\psi_1|\psi_2\rangle \sim 1$ for the overlap factor. With amplification (under the same conditions which gave a number gain of 30 (Fig. 2)), the visibility increased by a factor of about two over a

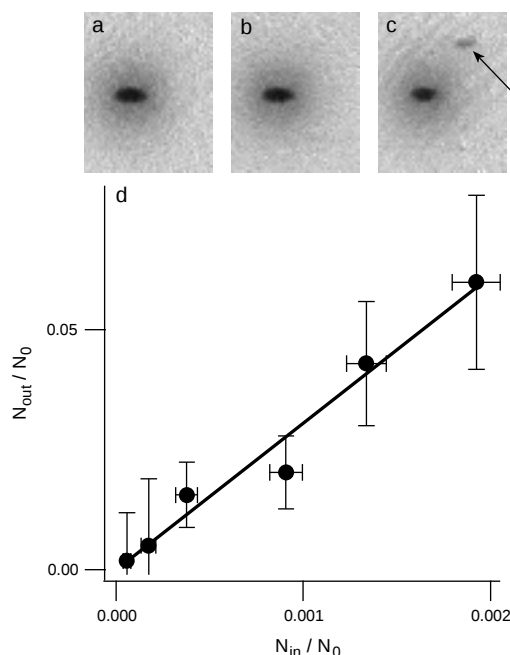


Figure 2 Input–output characteristics of the matter-wave amplifier. **a–c**, Typical time-of-flight absorption images demonstrating matter-wave amplification. The output of the seeded amplifier (**c**) is clearly visible, as indicated by the arrow, whereas no recoiling atoms are discernible in the case without amplification (**a**) or amplification without the input (**b**). The size of the images is 2.8 mm $\times$ 2.3 mm. **d**, Output of the amplifier as a function of the number of atoms at the input. A straight-line fit shows a number gain of 30. Note that the same data set as in Fig. 4 was used. The total number of amplified atoms was determined by measuring the increase in the offset of the obtained interference fringes discussed below.

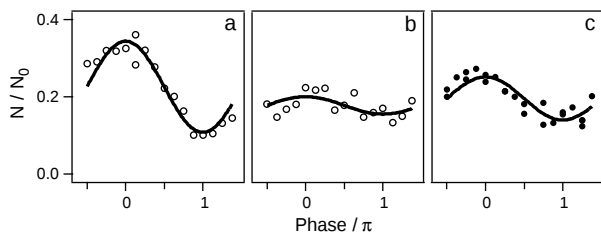


Figure 3 Phase-coherent amplification of matter waves. Shown is the number of atoms N in the recoil mode (normalized by the initial number of atoms in the condensate N_0) versus the relative phase between the two Bragg pulses. **a**, Interference between two matter waves of almost equal intensity yielded a high contrast fringe (visibility $V = 52\%$). **b**, When the intensity of the first matter wave was decreased by a factor of 40, the visibility V dropped to the noise level ($< 15\%$). **c**, By switching on the amplifier, a large visibility was regained ($V = 31\%$).

wide range of input intensities. Since for a small number of atoms N_1 , the visibility increases as $\sqrt{N_1}$ (equation (4)), this implies a “coherent matter-wave gain” of at least four. The discrepancy from the number gain of 30 may be due to a distortion of the input matter wave during its amplification which would reduce the overlap factor $\langle \psi_1 | \psi_2 \rangle$. Indeed, the shape of the amplified matter wave looked distorted after the 35 ms of ballistic expansion. This indicates a wavefront distortion of the amplified matter wave which can be parametrized by introducing an overlap factor of $\sqrt{4/30} = 0.4$ in equation (3), but this effect requires further study. No distortion was seen at short times (1.5 ms) after the amplification (as observed by phase-contrast imaging). This and a simple model suggest that the spatially non-uniform gain is not the dominant cause of the non-ideal visibility.

The interference between the (amplified) signal and the reference matter wave was measured in a two-pulse interferometer where the delay time between the pulses was kept shorter than the coherence time of the two-photon Bragg process ($\sim 40 \mu\text{s}$; ref. 7). The coherence time can be determined by the linewidth of the Bragg resonance¹¹, or by the decay rate of the Ramsey fringe visibility¹². In the absence of mean-field broadening, it is proportional to the spatial extent of the condensate divided by the recoil velocity^{11,12}.

We deliberately studied the matter-wave amplification in the small-signal limit. We chose a low density of the condensate, a small number of input atoms, and rather low gain in order to avoid additional complexities such as elastic collisions and repeated superradiant scattering⁷. The density of the condensate was adjusted to be less than $2 \times 10^{14} \text{ cm}^{-3}$ by limiting the number of condensed atoms. The average number of (incoherent) elastic collisions occurring between the two Bragg pulses was less than 0.05 per recoiling atom, and should be negligible.

The input matter waves were generated in perfect spatial overlap with the condensate by Bragg pulses. In further experiments, we have also observed the single-pass amplification of input atomic wave packets which were initially spatially separated from the condensate. This was achieved by letting the input wave packet oscillate in the trapping potential for half an oscillation period before applying the amplification light pulse. As the wave packet passed through the pumped condensate at rest, amplification was clearly observed.

The matter-wave amplifier that we have demonstrated amplifies only the recoil states that can be populated by normal Rayleigh scattering. Their momentum is located on a sphere with radius $|\hbar \mathbf{k}_0|$ centred at the incoming photon momentum ($\hbar \mathbf{k}_0$). We have achieved perfect momentum matching by using the radial beam for the Bragg process as the pump beam for the amplification process. Ultimately, the momentum bandwidth of the amplification process should be given by the momentum uncertainty of the condensate, which is roughly Planck’s constant divided by the size of the condensate.

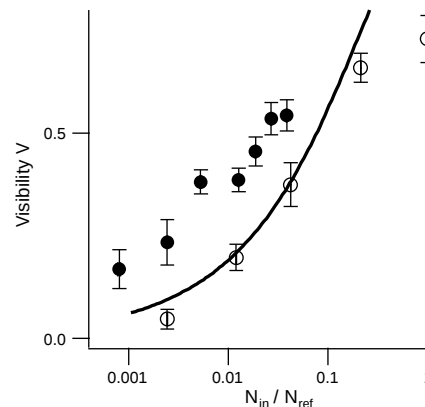


Figure 4 Visibility of interference with and without amplification. The visibility of the interference pattern between input (output) and the reference matter wave is shown with open (filled) circles as a function of the input number of atoms N_{in} (normalized by the population of the reference matter wave N_{ref}). The solid line is the expected visibility from equation (3), assuming an overlap factor of unity. The intensity of the amplification pulse was 40 mW cm^{-2} .

The amplification process studied in this work is based on bosonic stimulation: if N bosons occupy a given state, the transition rates into that state are proportional to $N + 1$. Some experiments have seen such accelerated transition rates without observing phase coherence^{7,13,14}. A study of the formation of the condensate showed evidence for bosonic stimulation in the elastic scattering of atoms into the condensate¹³, an irreversible process establishing thermal equilibrium. The recently demonstrated four-wave mixing of atoms¹⁴ also relied on bosonic stimulation of atom–atom collisions, but in this case as a coherent process mediated by the mean field.

Bosonic stimulation has been discussed as the gain mechanism of atom lasers, either based on evaporation or optical pumping^{1–4}. In these discussions, the idea was to start with a thermal cloud of low phase-space density and create a large population in a single matter-wave mode. Our experiments exploited the high phase-space density of a condensate to realize a high matter-wave gain. In principle, we could have used an optically pumped thermal cloud as a gain medium, but the gain would have been extremely small due to the large Doppler broadening.

Our experiment can also be regarded as a demonstration of an active atom interferometer. It realizes a two-pulse atom interferometer¹⁵ with phase-coherent amplification in one of the arms. Such active interferometers may be advantageous for precise measurements of phase shifts in highly absorptive media; for example, measurements of the index of (matter wave) refraction when a condensate passes through a gas of atoms or molecules¹⁶. As the most accurate optical gyroscopes involve active interferometers¹⁷, atom amplification might also play a role in future matter-wave gyroscopes¹⁸.

Note added in proof: M. Kozuma has informed us that related work has been done at the University of Tokyo. □

Received 9 November; accepted 18 November 1999.

- Holland, M., Burnett, K., Gardiner, C., Cirac, J. & Zoller, P. Theory of an atom laser. *Phys. Rev. A* **54**, R1757–R1760 (1996).
- Olshanii, M., Castin, Y. & Dalibard, J. in *Proc. XII Conf. on Laser Spectroscopy* (eds Inguscio, M., Allegrini, M. & Sasso, A.) 7–12 (World Scientific, New York, 1995).
- Spreeuw, R. J. C., Pfau, T., Janicke, U. & Wilkens, M. Laser-like scheme for atomic-matter waves. *Eur. Phys. Lett.* **32**, 469–474 (1995).
- Wiseman, H. M. & Collett, M. J. An atom laser based on dark-state cooling. *Phys. Lett. A* **202**, 246–252 (1995).
- Law, C. K. & Bigelow, N. P. Amplifying an atomic wave signal using a Bose-Einstein condensate. *Phys. Rev. A* **58**, 4791–4795 (1998).
- Moore, M. G. & Meystre, P. Optical control and entanglement of atomic Schrödinger fields. *Phys. Rev. A* **59**, R1754–R1757 (1999).
- Inoué, S. *et al.* Superradiant Rayleigh scattering from a Bose-Einstein condensate. *Science* **285**, 571–574 (1999).

8. Moore, M. G. & Meystre, P. Theory of superradiant scattering of laser light from Bose-Einstein condensates. Preprint cond-mat/9905425 at (<http://xxx.lanl.gov>) (1999).
9. Mewes, M.-O. *et al.* Bose-Einstein condensation in a tightly confining dc magnetic trap. *Phys. Rev. Lett.* **77**, 416–419 (1996).
10. Kozuma, M. *et al.* Coherent splitting of Bose-Einstein condensed atoms with optically induced Bragg diffraction. *Phys. Rev. Lett.* **82**, 871–875 (1999).
11. Stenger, J. *et al.* Bragg spectroscopy of a Bose-Einstein condensate. *Phys. Rev. Lett.* **82**, 4569–4573 (1999).
12. Hagley, E. W. *et al.* Measurement of the coherence of a Bose-Einstein condensate. *Phys. Rev. Lett.* **83**, 3112–3115 (1999).
13. Miesner, H.-J. *et al.* Bosonic stimulation in the formation of a Bose-Einstein condensate. *Science* **279**, 1005–1007 (1998).
14. Deng, L. *et al.* Four-wave mixing with matter waves. *Nature* **398**, 218–220 (1999).
15. Berman, P. R. (ed.) *Atom Interferometry* (Academic, New York, 1997).
16. Schmiedmayer, J. *et al.* Index of refraction of various gases for sodium matter waves. *Phys. Rev. Lett.* **74**, 1043–1047 (1995).
17. Stedman, G. E. *et al.* Ring-laser tests of fundamental physics and geophysics. *Rep. Prog. Phys.* **60**, 615–688 (1997).
18. Gustavson, T. L., Bouyer, P. & Kasevich, M. A. Precision rotation measurements with an atom interferometer gyroscope. *Phys. Rev. Lett.* **78**, 2046–2049 (1997).

Acknowledgements

We thank D. M. Stamper-Kurn for discussions. This work was supported by the Office of Naval Research, NSF, Joint Services Electronics Program, ARO, NASA, and the David and Lucile Packard Foundation. A.P.C. was supported by NSF, A.G. by DAAD, and T.P. by the Alexander von Humboldt Foundation.

Correspondence and requests for materials should be addressed to S.I. (e-mail: sinouye@mit.edu).

Weaker Gulf Stream in the Florida Straits during the Last Glacial Maximum

Jean Lynch-Stieglitz^{*}, William B. Curry[†] & Niall Slowey[‡]

^{*} Department of Earth and Environmental Sciences, Lamont-Doherty Earth Observatory, Palisades, New York 10964, USA

[†] Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

[‡] Department of Oceanography, Texas A & M University, College Station, Texas 77843, USA

As it passes through the Florida Straits, the Gulf Stream consists of two main components: the western boundary flow of the wind-driven subtropical gyre and the northward-flowing surface and intermediate waters which are part of the 'global conveyor belt', compensating for the deep water that is exported from the North Atlantic Ocean¹. The mean flow through the Straits is largely in geostrophic balance and is thus reflected in the contrast in seawater density across the Straits². Here we use oxygen-isotope ratios of benthic foraminifera which lived along the ocean margins on the boundaries of the Florida Current during the Last Glacial Maximum to determine the density structure in the water and thereby reconstruct transport through the Straits using the geostrophic method—a technique which has been used successfully for estimating present-day flow³. Our data suggest that during the Last Glacial Maximum, the density contrast across the Florida Straits was reduced, with the geostrophic flow, referenced to the bottom of the channel, at only about two-thirds of the modern value. If the wind-driven western boundary flow was not lower during the Last Glacial Maximum than today, these results indicate a significantly weaker conveyor-belt component of the Gulf Stream compared to present-day values. Whereas previous studies based on tracers suggested that deep waters of North Atlantic origin were not widespread during glacial times, indicating either a relatively weak or a shallow overturning cell, our results provide evidence that the overturning cell was indeed weaker during glacial times.

The Florida Current, which flows through the Florida Straits (Fig. 1), is the southernmost part of the Gulf Stream. The average northward transport of the Florida Current is fairly well constrained by modern measurement at 30–32 Sv, and shows a seasonal variation with a range of 4.6 Sv as well as considerable variability on shorter timescales⁴. Schmitz and McCartney¹ assign an uncertainty of only 5% to the mean annual transport value of 31 Sv. The transport at the Florida Current includes 13 Sv of flow from the South Atlantic which travels northward in the Gulf Stream to the North Atlantic and ultimately compensates for the export of North Atlantic Deep Water (NADW). The other 17 Sv compensate for southward-flowing upper waters from the eastern portion of the wind-driven North-Atlantic subtropical gyre¹. The strong tilts in the surfaces of constant temperature and density within the Florida Straits reflect the geostrophic adjustment of the density surfaces in the presence of the large velocities (Fig. 2a and b). The contrast in temperature and density across the Florida Current is large, and is well represented in the $\delta^{18}\text{O}$ of benthic foraminifera living in this region.

We can use the $\delta^{18}\text{O}$ from the calcite tests of foraminifera to estimate density because both the $\delta^{18}\text{O}$ of calcite ($\delta^{18}\text{O}_{\text{calcite}}$) and density increase as a result of increasing salinity or decreasing temperature. The dependence of seawater density on salinity and temperature is well known and will be constant throughout the ocean and through geological time. The dependence of $\delta^{18}\text{O}$ of foraminifera on the temperature of calcification is also fairly well constrained. The fractionation between calcite precipitated inorganically and the water in which it forms increases by about 0.2‰ for every 1 °C decrease in temperature⁵, and the isotopic composition of calcitic benthic foraminifera in the genera *Planulina* and *Cibicides* show the same fractionation as measured in the experiments with inorganic calcite³. The relationship between $\delta^{18}\text{O}_{\text{calcite}}$ and salinity is more complex. The $\delta^{18}\text{O}_{\text{calcite}}$ reflects the $\delta^{18}\text{O}$ of the water in which the foraminifera grew. The $\delta^{18}\text{O}$ of sea water ($\delta^{18}\text{O}_{\text{water}}$) primarily reflects patterns of evaporation and freshwater influx to the surface of the ocean. Because salinity also reflects these patterns, salinity and $\delta^{18}\text{O}_{\text{water}}$ are often well correlated in the ocean. Although the exact relationship varies in different areas of the surface ocean⁶, the vast majority of surface and warm subsurface waters ($T > 5^\circ\text{C}$) in the ocean have salinity and $\delta^{18}\text{O}_{\text{water}}$ values which scatter around a linear trend³. For times in the geological past, our ability to reconstruct density from the $\delta^{18}\text{O}_{\text{calcite}}$

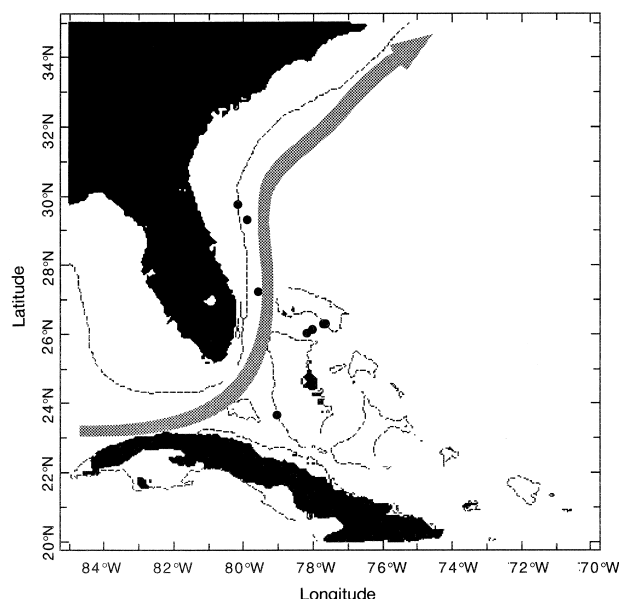


Figure 1 Location of sediment cores used in this study. The path of the Gulf Stream is indicated with the shaded arrow, and the depth of the 200-m isobath with a dashed line.

is most limited by our knowledge of the relationships between $\delta^{18}\text{O}_{\text{water}}$ and salinity, as well as the relationships between temperature and salinity. However, while the patterns of atmospheric and surface ocean circulation might have shifted in latitude, we still expect that there existed an evaporative subtropical gyre and a fresh sub-polar gyre even during the periods of maximum glaciation. By taking into consideration the presumed changes in ocean salinity and $\delta^{18}\text{O}_{\text{water}}$ due to sea-level change, and possible changes in the

$\delta^{18}\text{O}$ of high-latitude precipitation, we can make an educated guess about the relationship between $\delta^{18}\text{O}_{\text{water}}$ and salinity for the glacial ocean. We can then determine the relationship between $\delta^{18}\text{O}_{\text{calcite}}$ and density for the glacial ocean (see Methods).

In order to reconstruct vertical profiles of $\delta^{18}\text{O}_{\text{calcite}}$, which can then be converted to density, we measured the $\delta^{18}\text{O}$ in individual benthic foraminifera (genera *Planulina* and *Cibicides*) from sediment cores on either side of the Florida Straits (Fig. 1). The benthic foraminifera data from four of the cores from the seaward (Bahamas) side of the current are from a previous study⁷. The isotopic data from the remaining cores used in this study are shown in Fig. 3. We used grouped $\delta^{18}\text{O}$ measurements from a planktonic foraminifera, *G. sacculifer*, to identify the Holocene and Last Glacial Maximum (LGM) levels in the cores and to constrain the surface ocean values of $\delta^{18}\text{O}_{\text{calcite}}$. The inferred distribution of $\delta^{18}\text{O}_{\text{calcite}}$ across the Florida Straits reconstructed from the core tops reflects the modern density structure in this region (Fig. 2). During the LGM there was very little contrast in $\delta^{18}\text{O}_{\text{calcite}}$ across the deeper portion of the Straits. If the LGM Florida Current was, like today's, largely baroclinic, this would imply very little northward flow at these depths. This is consistent with previous work which suggests little northward penetration of Antarctic Intermediate Water into the Caribbean during the LGM^{8,9}. The tilting lines of constant $\delta^{18}\text{O}_{\text{calcite}}$ in the upper part of the section suggest that there was northward flow of surface waters through the glacial Florida Straits. To use the geostrophic method to calculate the flow from the vertical profiles of $\delta^{18}\text{O}_{\text{calcite}}$ (Fig. 4), we assume that the transport decreases to zero near the bottom of the channel, as is observed today. The core top profiles yield a transport of 30 Sv, in line with an estimate using a more extensive set of surface sediment data (32 Sv) (ref. 3) and modern estimates from physical oceanographic studies (30–32 Sv). Using our best estimate for the relationship between glacial $\delta^{18}\text{O}_{\text{calcite}}$ and density, the LGM profiles result in a transport estimate of 15–18 Sv. This reduced flow does not result simply from raising the reference level from 760 m (the depth of the channel today) to 640 m (the depth of the LGM channel). If we calculated the modern transport using a reference level of 640 m instead of 760 m, the flow is only reduced from 30 Sv to 27 Sv. By varying the assumptions that went into our estimate of the relationship between glacial $\delta^{18}\text{O}_{\text{calcite}}$ and density within reasonable bounds, the LGM transport could have been as low as 14 Sv or as high as 21 Sv (see Methods). While today's flow is largely baroclinic we cannot rule out the possibility of a barotropic component to the flow during the last ice age. However, given that there is currently no evidence otherwise, we presume that, like today, the flow is small at the sill depth of the Florida Straits.

What could have lowered the transport through the Florida Straits during the LGM? While sedimentological evidence suggests increased aridity for the last glaciation, there is little direct evidence addressing wind strength in the glacial North Atlantic. However, atmospheric general circulation models driven by glacial boundary conditions suggest that the winds which drive the gyre circulation were either unchanged or slightly strengthened relative to today^{10,11}. Stronger winds would have caused the wind-driven flow of upper waters through the Florida Straits to increase. A simple shift of the wind belts towards the Equator would also have caused an increase in the wind-driven transport. However, we cannot exclude the possibility of more complex changes in wind patterns which decrease the wind stress curl and, thus, the wind-driven Sverdrup transport at the latitude of the northernmost passage into the Caribbean. We believe that the simplest explanation for the reduction in the Florida Current is the reduction or absence of the interhemispheric component of the transport (13 of the 30 Sv flowing through the Florida Straits today), which compensates deep water formation in the North Atlantic.

There is substantial evidence that during the LGM, NADW did not dominate the deep Atlantic as it does today^{12–14}. If there were no

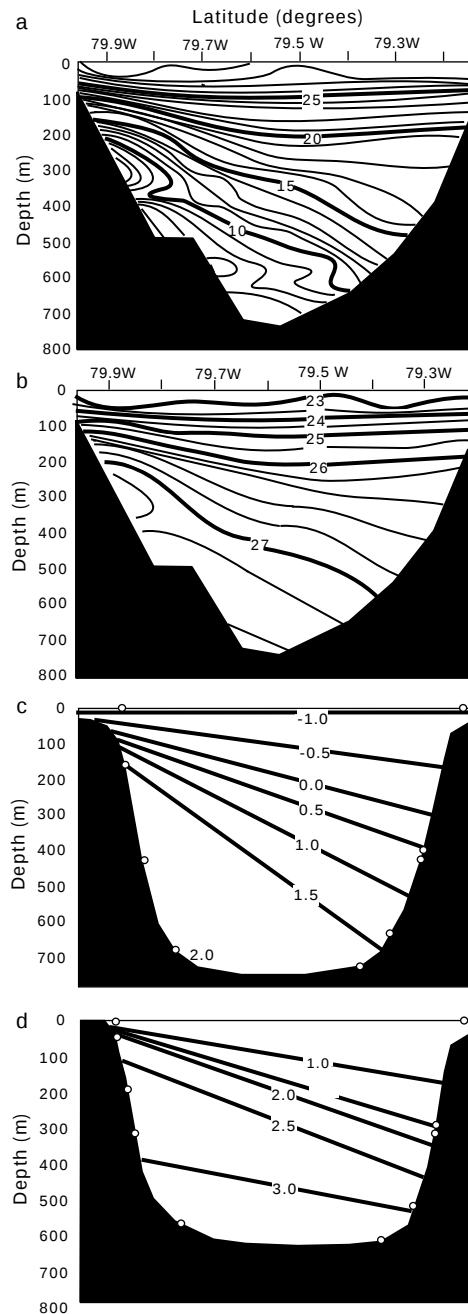


Figure 2 Observed and inferred hydrographic sections across the Florida Straits. **a, b**, Temperature in °C (**a**), density (σ_t) (**b**) observed across the Florida Straits at 27 N. Data are from the World Ocean Atlas CTD data set²⁹. **c, d**, Section of $\delta^{18}\text{O}_{\text{calcite}}$ at a similar location for the Holocene (**c**) and LGM (**d**) oceans is inferred from the foraminiferal $\delta^{18}\text{O}_{\text{calcite}}$ shown in Figs 3 and 4. Black dots indicate the depths of the $\delta^{18}\text{O}_{\text{calcite}}$ measurements used in calculating the geostrophic flow. The LGM bathymetry has been adjusted upwards by 120 m to account for the sea level drop²⁶. The Holocene $\delta^{18}\text{O}_{\text{calcite}}$ reconstruction (**c**) reflects the observed density structure. The LGM section (**d**) suggests the concentration of flow at shallower depths.

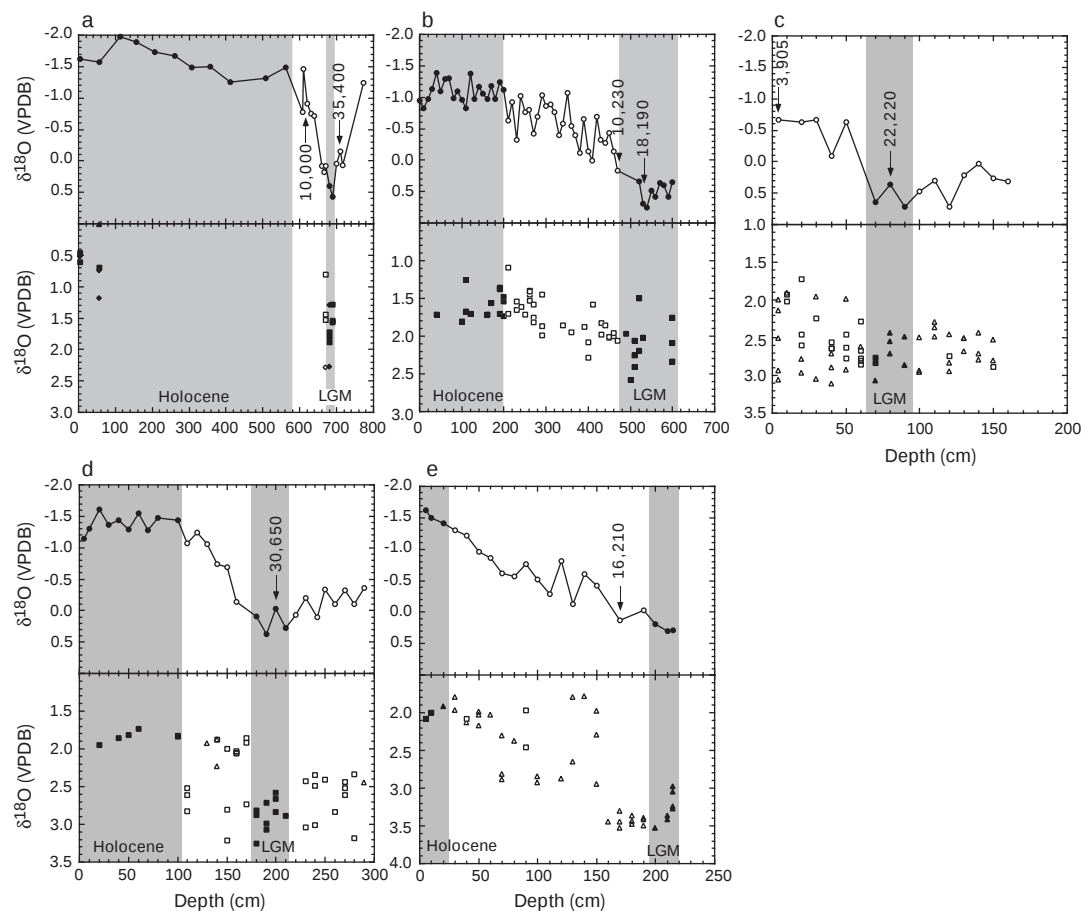


Figure 3 Isotopic measurements from sediment cores in this study. Oxygen-isotope ratios measured on the planktonic foraminifera *G. sacculifer* (circles) and individual benthic foraminifera, *P. ariminensis* (triangles), *C. pachyderma* (squares) and *C. foveolata* (diamonds). Filled symbols represent data contributing to LGM and Holocene averages. Radiocarbon dates (reservoir corrected) are also indicated at the corresponding depth in

the cores. **a**, ODP Site 1008 at the Bahamas, 437 m. Planktonic foraminifera data and radiocarbon are from ref. 31. **b**, Florida slope core RC1-1, 174 m. **c**, Florida slope core RC1-2, 324 m. **d**, Florida slope core V7-13, 452 m. **e**, Florida slope core V3-149, 706 m. LGM, last glacial maximum.

deep-water formation in the North Atlantic, then the compensating northward surface-water transport would not have been present, explaining the observed decrease in transport through the Florida Straits. However, while NADW production appears to have decreased, the presence of a nutrient-depleted water mass at shallower depths (<2,000 m)^{14,15} suggests a compensating increase in intermediate water export from the glacial Atlantic. Two- and three-dimensional ocean-circulation models do show a shift from deep to intermediate water production in the glacial North Atlantic under ice-age conditions (increased high-latitude freshwater flux)^{16,17}. In these models, the production of Glacial North Atlantic Intermediate Water (GNAIW) is, like today's production of NADW, compensated by a northward flow of warm surface waters. The reduced western boundary transport we observe at the Florida Straits would require that the levels of nutrient depletion in the intermediate waters of the glacial Atlantic are achieved with a very weak overturning circulation. Sigman and Lehman¹⁸ suggest that as little as 5 Sv of GNAIW formation are needed to explain the low Cd/Ca and high $\delta^{13}\text{C}$ observed in the glacial North Atlantic, as long as a significant fraction of this water is exported from the Atlantic. The export of GNAIW is supported by radiochemical¹⁹ as well as nutrient-linked water mass tracers²⁰. Other modelling studies²¹ also suggest that a reduced meridional overturning circulation is consistent with LGM tracer distributions.

Alternatively, GNAIW may have been ventilated by a different mechanism altogether, one which did not involve the compensating

northward flow of warm surface waters. Planktonic foraminifera assemblage data suggest a nearly zonal polar front separating the subtropics and subpolar gyres for the LGM North Atlantic²². This is consistent with the absence of a glacial 'North Atlantic Drift' bringing surface waters into high northern latitudes, and consequently, the presence of a well developed subpolar gyre. In addition, oxygen-isotope measurements in planktonic and benthic foraminifera suggest that surface waters in the region of the presumed LGM subpolar gyre are the only waters which could become dense enough to form the GNAIW²³. If GNAIW were ventilated by deep convection in the subpolar gyre, the southward-flowing GNAIW would have been replaced primarily by northward-flowing intermediate waters which would have been too dense to have passed through the relatively shallow glacial Florida Straits.

The northward heat transport by the meridional overturning circulation is proportional to both the strength of the overturning circulation, and the temperature difference between the northward- and southward-flowing components. The very weak overturning circulation required if GNAIW was compensated by northward-flowing surface waters would result in a northward heat transport greatly reduced relative to today. Similarly, any scenario for the meridional circulation with cold intermediate waters compensating GNAIW export implies significantly less oceanic heat transport into the North Atlantic than today, and also less heat transport than predicted by the ocean models with a shallow but still strong conveyor circulation in the glacial Atlantic. □

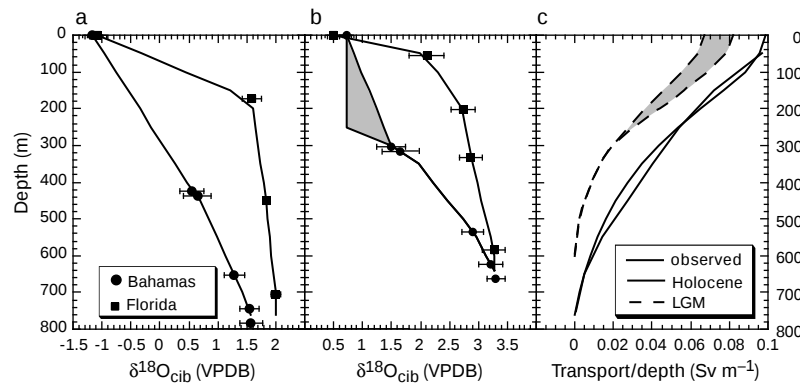


Figure 4 Modern and LGM oxygen isotope and transport profiles. **a**, **b**, Holocene (**a**) and LGM (**b**) $\delta^{18}\text{O}_{\text{calcite}}$ profiles. Error bars represent the 1σ standard deviation of the population of the $\delta^{18}\text{O}$ on individual foraminifera that went into the average for each core and level (Fig. 3). For the seaward (Bahamas) profile we show both a linear interpolation (minimum flow estimate) between the shallowest benthic foraminiferal data point and the surface, and an interpolation with a very deep mixed layer (maximum flow estimate).

c, The vertical distribution of transports calculated using our best estimates of the relationship between density and $\delta^{18}\text{O}_{\text{calcite}}$. The total transport is 30 Sv for the Holocene (black line) and 16–19 Sv for the LGM (dashed lines). The vertical structure of today's transport through the Florida Straits at 27°N as determined by current profilers is also shown (grey line)³⁰.

Methods

Measurements on *G. sacculifer* for the cores shown in Fig. 3 were made at the Lamont-Doherty Earth Observatory on a Micromass Optima with Multiprep, except for ODP Site 1008 where measurements were made at Texas A&M University (ref. 31). *G. sacculifer* were picked from the 355–425- μm size fraction, and groups of 5–8 individuals were analysed. Measurements on individual benthic foraminifera (*P. ariminensis*, *C. pachyderma* and *C. foveolata*) were made at Woods Hole Oceanographic Institution using the methods described by Curry²⁴. All oxygen-isotope measurements were converted to VPDB (Vienna Pee Dee Belemnite) via NBS-19. Radiocarbon dates for the Florida Slope cores are from planktonic foraminifera and those for ODP Site 1008 are from bulk sediments; all radiocarbon dates are corrected for a 400-year reservoir effect.

Vertical profiles of $\delta^{18}\text{O}_{\text{calcite}}$ used to calculate geostrophic transport for the Holocene and the LGM are constructed as follows. The modern and LGM intervals for the sediment cores are identified based on the $\delta^{18}\text{O}$ stratigraphy of grouped analyses of the surface-dwelling planktonic foraminifera, *G. sacculifer*. Planktonic foraminiferal (*G. sacculifer*) $\delta^{18}\text{O}$ data from V7-13 (Fig. 3d) and 103GGC (Little Bahama Banks)²⁵ are used for the sea-surface data point in each profile. Average $\delta^{18}\text{O}_{\text{calcite}}$ for the Holocene and LGM sections of each core is computed from the measurements on the individual benthic foraminifera. Holocene and LGM averages from Slowey and Curry⁷ are used for the deeper Bahamas cores. The core depths for the LGM data have been adjusted by 120 m to reflect the lowered sea level²⁶. For the landward (Florida) profile, we linearly interpolate between the data at the core depths. For the seaward (Bahamas) profile, two different interpolations between the shallowest benthic foraminiferal data point and the surface are used, a linear interpolation (minimum flow estimate), and an interpolation with a very deep mixed layer (maximum flow estimate).

We then convert the Holocene $\delta^{18}\text{O}_{\text{calcite}}$ to density using the relationship for the modern ocean ($T > 5^\circ\text{C}$) described previously³. For the LGM, we assume that the relationship between temperature and salinity in the glacial ocean was similar to that of today, only that salinity was concentrated by a factor of 1.03 (about 1 psu) due to the build-up of continental ice. In addition, we assume that the mean ocean $\delta^{18}\text{O}$ increased by 1‰ for the same reasons²⁷. We also assume that the relationship between $\delta^{18}\text{O}$ of sea water and salinity represents a mixing between sub-thermocline water and a fresh end member that has a salinity of zero and a $\delta^{18}\text{O}$ representative of precipitation in subpolar latitudes⁶. A study using an atmospheric general circulation model²⁸ suggests that during the LGM this subpolar precipitation was about 4‰ lower than today. We then propose that for warm waters in the glacial ocean $\delta^{18}\text{O}$ was related to salinity as follows: $\delta^{18}\text{O} = -19.3 + 0.52 \times \text{salinity}$. These changes result in a relationship between $\delta^{18}\text{O}_{\text{calcite}}$ and density, σ_t , for the glacial ocean of: $\sigma_t = 25.4 + 1.3\delta^{18}\text{O}_{\text{calcite}} - 0.12\delta^{18}\text{O}_{\text{calcite}}^2$. The geostrophic transport between the two profiles was calculated using the method described previously³. For the core tops, we assign a level of no motion at a depth of 760 metres, the depth of the strait at its most constricted point. For the LGM we adjust the level of no motion upward by 120 metres (to 640 metres) to account for the lowered sea surface. We obtain a transport of 30 Sv for the Holocene and 15–18 Sv for the LGM, using the relationship between density and $\delta^{18}\text{O}_{\text{calcite}}$ derived above, and the two different interpolated profiles for the seaward side of the Straits. We vary the assumed LGM mean ocean $\delta^{18}\text{O}$ change between 1 and 1.3‰ (ref. 26) higher than present, and the assumed LGM change in the $\delta^{18}\text{O}$ of high-latitude precipitation between 0‰ and 8‰ lower than present. A maximum transport of 21 Sv is obtained with a mean ocean $\delta^{18}\text{O}$ change of 1.3‰ and a $\delta^{18}\text{O}$ of high-latitude precipitation 8‰ lower than present. A minimum transport of 14 Sv is obtained with a mean ocean $\delta^{18}\text{O}$ change of 1‰ and a $\delta^{18}\text{O}$ of high-latitude precipitation the same as at present.

Received 4 May; accepted 12 October 1999.

1. Schmitz, W. J. & McCartney, M. S. On the North Atlantic Circulation. *Rev. Geophys.* **31**, 29–49 (1993).

- Wust, G. Florida und Antillenstrom: Eine hydrodynamische Untersuchung. *Geographischnaturwiss. Reihe* **12**, 48 (1924).
- Lynch-Stieglitz, J., Curry, W. & Slowey, N. A geostrophic transport estimate for the Florida Current from oxygen isotope composition of benthic foraminifera. *Paleoceanography* **14**, 360–373 (1999).
- Larsen, J. C. Transport and heat-flux of the Florida Current at 27-degrees-N derived from cross-stream voltages and profiling data—Theory and observations. *Phil. Trans. R. Soc. Lond. A* **338**, 169–236 (1992).
- Kim, S. T. & O'Neil, J. R. Equilibrium and nonequilibrium oxygen isotope effects in synthetic carbonates. *Geochim. Cosmochim. Acta* **61**, 3461–3475 (1997).
- Craig, H. & Gordon, L. I. in *Pro. 3rd Spoleto Conf., Spoleto, Italy* (ed. Tongiorgi, E.) 9–130 (Sischi and Figli, Pisa, 1965).
- Slowey, N. C. & Curry, W. B. Glacial-interglacial differences in circulation and carbon cycling within the upper western North-Atlantic. *Paleoceanography* **10**, 715–732 (1995).
- Haddad, G. A. & Droxler, A. W. Metastable CaCO_3 dissolution at intermediate water depths of the Caribbean and western North Atlantic: Implications for intermediate water circulation during the past 200,000 years. *Paleoceanography* **11**, 701–716 (1996).
- Marchitto, T. M., Curry, W. B. & Oppo, D. W. Millennial-scale changes in North Atlantic circulation since the last glaciation. *Nature* **393**, 557–561 (1998).
- Kutzbach, J. et al. Climate and biome simulations for the past 21,000 years. *Quat. Sci. Rev.* **17**, 473–506 (1998).
- Seidov, D., Sarinthein, M., Stratteger, K., Prien, R. & Weinelt, M. North Atlantic ocean circulation during the last glacial maximum and subsequent meltwater event: A numerical model. *J. Geophys. Res.* **101**, 16305–16332 (1996).
- Boyle, E. A. & Keigwin, L. D. North Atlantic thermohaline circulation during the last 20,000 years linked to high latitude surface temperature. *Nature* **330**, 35–40 (1987).
- Duplessy, J.-C. et al. Deepwater source variation during the last climatic cycle and their impact on global deepwater circulation. *Paleoceanography* **3**, 343–360 (1988).
- Sarinthein, M. et al. Changes in east Atlantic deep-water circulation over the last 30,000 years—8 time slice reconstructions. *Paleoceanography* **9**, 209–267 (1994).
- Oppo, D. W. & Lehman, S. J. Mid-depth circulation of the subpolar North-Atlantic during the Last Glacial Maximum. *Science* **259**, 1148–1152 (1993).
- Fichefet, T., Hovine, S. & Duplessy, J.-C. A model study of the Atlantic thermohaline circulation during the last glacial maximum. *Nature* **372**, 252–255 (1994).
- Ganopolski, A., Rahmstorf, S., Petoukhov, V. & Claussen, M. Stimulation of modern and glacial climates with a coupled global model of intermediate complexity. *Nature* **391**, 351–356 (1998).
- Sigman, D. M. & Lehman, S. J. in *American Geophysical Union Fall Meeting (AGU)*, San Francisco, California, 1995).
- Yu, E. F., Francois, R. & Bacon, M. P. Similar rates of modern and last-glacial ocean thermohaline circulation inferred from radiochemical data. *Nature* **379**, 689–694 (1996).
- Lynch-Stieglitz, J., vanGeen, A. & Fairbanks, R. G. Inter-ocean exchange of Glacial North Atlantic intermediate water: Evidence from Subantarctic Cd/Ca and carbon isotope measurements. *Paleoceanography* **11**, 191–201 (1996).
- Winguth, A. M. E., Archer, D., Duplessy, J.-C., Maier-Reimer, E. & Mikolajewicz, U. Sensitivity of paleonutrient tracer distributions and deep-sea circulation to glacial boundary conditions. *Paleoceanography* **14**, 304–323 (1999).
- McIntyre, A. et al. in *Investigation of Late Quaternary Paleo-Oceanography and Paleoclimatology* (eds Cline, R. M. & Hays, J. D.) 43–76 (The Geological Society of America, Boulder, 1976).
- Labeyrie, L. D. et al. Changes in the vertical structure of the North-Atlantic Ocean between glacial and modern times. *Quat. Sci. Rev.* **11**, 401–413 (1992).
- Curry, W. B. in *The South Atlantic: Present and Past Circulation* (eds Wefer, G., Berger, W. H., Siedler, G. & Webb, D.) (Springer, New York, 1996).
- Slowey, N. C. & Curry, W. B. Enhanced ventilation of the North Atlantic subtropical gyre thermocline during the last glaciation. *Nature* **358**, 665–668 (1992).
- Fairbanks, R. G. A 17,000-year glacio-eustatic sea-level record—Influence of glacial melting rates on the Younger Dryas Event and deep-ocean circulation. *Nature* **342**, 637–642 (1989).
- Schrag, D. P., Hampt, G. & Murray, D. W. Pore fluid constraints on the temperature and oxygen isotopic composition of the glacial ocean. *Science* **272**, 1930–1932 (1996).

28. Joussaume, S. & Jouzel, J. Paleoclimatic tracers—An investigation using an atmospheric general-circulation model under ice-age conditions. 2. Water isotopes. *J. Geophys. Res. Atmos.* **98**, 2807–2830 (1993).
29. Levitus, S. & Boyer, T. P. *World Ocean Atlas 1994* (National Oceanic and Atmospheric Administration, Washington DC, 1994).
30. Leaman, K. D., Johns, E. & Rossby, T. The average distribution of volume transport and potential vorticity with temperature at 3 sections across the Gulf-Stream. *J. Phys. Oceanogr.* **19**, 36–51 (1989).
31. Slowey, N. C. *et al.* Glacial to Holocene sedimentation on the western slope of Great Bahama Bank. *Geo-Mar. Lett.* (submitted).

Acknowledgements

This work was supported by grants from the US National Science Foundation and a grant/cooperative agreement from the National Oceanic and Atmospheric Administration. The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its subagencies. Support for the curating facilities of the Lamont-Doherty Earth Observatory Deep-Sea Sample Repository is provided by the National Science Foundation and the Office of Naval Research. We are grateful to J. Mayer, A. LeGrande, D. Ostermann and M. Yeager for technical assistance.

Correspondence and requests for materials should be addressed to J.L.-S. (e-mail: jean@ldeo.columbia.edu).

A criterion for the fragmentation of bubbly magma based on brittle failure theory

Youxue Zhang

Department of Geological Sciences, The University of Michigan, Ann Arbor, Michigan 48109-1063, USA

The fragmentation of bubbly magma is a defining point in a volcanic eruption—before fragmentation the magma flows relatively slowly, during fragmentation the bubbles break up to release compressed gas and, afterwards, the eruption becomes a violent gas flow carrying suspended magma particles. Seemingly benign lava flows or domes can suddenly fragment into deadly pyroclastic flows^{1–3}. Several criteria have been proposed to define the point of magma fragmentation or foam stability^{4–7}. The criterion of Papale⁷ is based on melt relaxation theory and equates magma strain rate with the rate of increase of flow velocity with distance. It ignores, however, the role of bubble pressure in causing fragmentation. Two empirical approaches^{4,5} consider the role of high bubble pressure in causing fragmentation but do not address the underlying physics of magma fragmentation. Here I develop a fragmentation criterion for bubbly magma based on brittle failure theory and apply it to the fragmentation of lava domes and flows. On the basis of this theory, a bubbly magma will fragment when the tensile stress at the inner walls of bubbles exceeds the tensile strength of the magma. The fragmentation conditions depend strongly on initial water content, with calculated vesicularity and final water levels coinciding reasonably well with those in observed pumices. This suggests that the proposed criterion captures the essence of the fragmentation process in bubbly magma.

Table 1 Calculated conditions for fragmentation

Given conditions				Fragmentation conditions	
<i>T</i> (°C)	<i>S</i> (bar)	<i>P</i> _{out} (bar)	H ₂ O _{t,i} (wt%)	Vesicularity (%)	H ₂ O _{t,av} (wt%)
700	60	1	1.0	57	0.66
700	60	3	1.0	63	0.63
700	50	1	1.0	47	0.74
700	50	1	0.7	77	0.35

H₂O_{t,av} is the average H₂O_i in the melt shell at the time of fragmentation.

Fragmentation can be viewed as the result of brittle failure of many bubbles at roughly the same time. Hartog⁸ summarized classic theories of material strength and concluded that (1) the maximum-strain theory has been discredited by experiments; (2) the maximum-stress theory applies well to the failure of brittle materials; and (3) the maximum-shear theory applies well to the beginning of yield in ductile materials. The maximum-stress theory states that brittle failure occurs when the maximum tensile stress ($\sigma_{\max}$) exceeds the tensile strength of the material (*S*):

$$\sigma_{\max} > S \quad (1)$$

In modern treatment of fracture mechanics, brittle failure occurs when the stress intensity factor exceeds the fracture toughness⁹. The stress intensity factor can be calculated, given the size and shape of any microcrack or visible crack. However, because the distribution, size, and shape of the initial microcracks and weaknesses are not known *a priori* in a magma, it is difficult to apply the modern approach. Nevertheless, the application of modern fracture mechanics arguments to a material containing numerous random small cracks leads to the modified Mohr theory of failure, which is equivalent to the maximum-stress theory if the sum of the greatest and least principal stresses is positive (J. R. Barber, personal communication). Because the maximum-stress theory is relatively easy to apply, it will be used in this work.

A bubbly magma system is complicated by the non-uniform distribution of bubbles of variable sizes. A first-order approximation is to assume that all bubbles are spherical, of the same size, and spaced regularly (Fig. 1a). Because the stress distribution and bubble growth rate in this case is still too complicated, it is further approximated by assuming that each spherical bubble is surrounded by a spherical shell (Fig. 1b), for which an analytical solution of stress distribution can be obtained and a numerical solution of bubble growth is available. Where the pressure at the outside surface of the shell is *P*_{out} and the pressure in the bubble is *P*_{in}, the stress at the bubble wall (inner wall of the shell) can be expressed as¹⁰:

$$\sigma^r = -P_{\text{in}} \quad (2)$$

$$\sigma^t = (P_{\text{in}} - P_{\text{out}}) \frac{1 + 2x}{2(1 - x)} - P_{\text{out}} \quad (3)$$

Here σ^r is the radial stress (and one of the principal stresses), σ^t is the tangential stress (and two of the principal stresses), and *x* is vesicularity. Tensile stress is positive and compressive stress is negative. (*P*_{in} − *P*_{out}) is referred to as ΔP hereafter and can be identified as the dynamic pressure, *P*_{dyn} (refs 11, 12). That ΔP is non-zero means that stress in the magma is not dissipated by viscous flow. That is, the fragmentation criterion for bubbly magma using the maximum-stress theory also automatically incorporates consideration of the liquid–glass transition.

When *P*_{in} > *P*_{out} > 0, then σ^r is compressive and σ^t can be tensile. The fracture criterion under compressive stresses is more complicated and is not considered here. When $\sigma^t + \sigma^r$ is positive, the maximum-stress theory applies. The maximum tensile stress is the tangential stress at the bubble wall (σ^t above). When the maximum tensile stress (σ^t) exceeds the tensile strength of the melt, the shells surrounding bubbles fail. That is, brittle failure occurs when the following holds:

$$\frac{1 + 2x}{2(1 - x)} \Delta P - P_{\text{out}} > S \quad (4)$$

Increasing *P*_{in} or increasing *x* (that is, decreasing the shell thickness) increases the tensile stress, and hence increases the likelihood of brittle failure. Because the crack volume is small and because *P*_{in} is exerted by the gas, the development of cracks does not significantly reduce ΔP and hence does not relieve the stress. Therefore, the cracks will grow until the bubble wall is broken (and the stress is thus relieved). This is in contrast to cracking

induced by thermal stress (cooling or heating). In the latter case, the development of cracks relieves the stress and hence crack growth may stop without completely breaking the material. In a bubbly magma, there is a range of bubble sizes and different bubbles are under different stress conditions. Bubbles with a narrow size distribution and under similar stress conditions break roughly simultaneously, leading to fragmentation.

When equation (4) is compared with the earlier empirical fragmentation criterion^{4,5}, $x\Delta P \geq S$, the difference is large. Although the earlier empirical approach does not apply quantitatively, it does capture the role of bubbles in fragmentation, unlike other work⁷.

Equation (1) above is general, and equation (4) represents a first-order approximation of the complicated real magma systems. In real volcanic systems, bubbles are of different sizes and distributed randomly (instead of uniformly); shells surrounding bubbles are not spherical, but consist of films separating two bubbles and plateau borders⁶ (Fig. 1); and the films have variable thicknesses. Because stress tends to concentrate at corners or thin films, all these complexities help to concentrate the stress and make the system fragment more easily. Furthermore, only part of the outside surface of a bubble is exposed to P_{out} and the rest is embedded in the magma body (Fig. 1a). Future work using equation (1) and incorporating these complexities (numerical solution of stress distribution) is necessary to produce more realistic fragmentation models.

The strength of magma has been investigated experimentally. Webb and Dingwell¹³ determined the tensile strength of fibre basaltic to rhyolitic melt to be 2500–5000 bar. Decrepitation experiments¹⁴ confirmed that dry glasses with CO₂ or Xe vesicles have very high strength, but H₂O-bearing glasses have much lower strengths of 15–55 bar (the strengths are calculated from equation (4) using data¹⁴ with the reported vesicularity; this is a correction to previous work¹⁴ using a zero vesicularity). Experiments have been carried out to investigate the fragmentation of the dacitic melt¹⁵ of Mount St Helens. For a 36% vesicularity of the melt, of which open vesicularity accounts for 29% and closed vesicularity accounts for only 7%, fragmentation has been shown to occur (Alidibirov *et al.*, personal communication) at $\Delta P \geq 30$ bar at 900 °C and ≥ 90 bar at 20 °C. Hence, using equation (4), $S \approx 40$ bar at 900 °C and $S \approx 120$ bar at 20 °C. By simple interpolation, $S \approx 60$ bar at 700 °C. However, because most bubbles in the experiments are interconnected, the gas in the bubbles escapes easily and may not exert much stress on bubble walls. Furthermore, closed bubbles were probably not

pressurized during the fragmentation experiments. Hence the obtained strength is a maximum. That is, actual strength is likely to be less than 60 bar at 700 °C.

Once the tensile strength of the magma is known, the fragmentation point can be determined from equation (4) by calculating P_{dyn} as a bubble grows in a shell of magma (Fig. 1) using a program¹². The accuracy of bubble growth calculation has been confirmed experimentally¹⁶ using newly assessed H₂O diffusivity and solubility¹⁷, provided a small correction is made to the viscosity model¹⁸ (a factor of 1.4–3.6, well within the stated uncertainty).

The fragmentation criterion developed above can be used to address volcanological problems. A long-standing puzzle for volcanologists is the sudden fragmentation of seemingly benign lava domes and flows into dangerous pyroclastic flows^{1–3}. To investigate the conditions for such fragmentation, I calculate bubble pressure and bubble wall stress using the method described above,

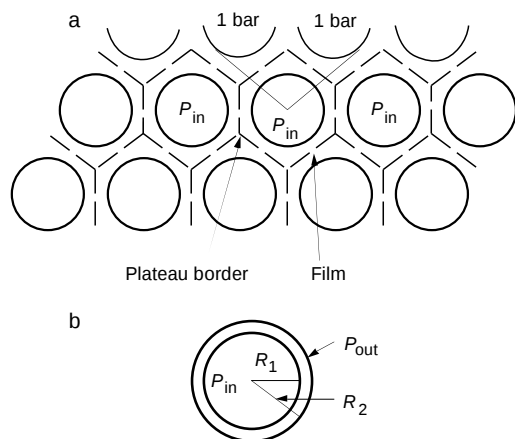


Figure 1 Simplification of real bubbly magma. **a**, By ignoring the variable bubble sizes, and non-uniform distribution, regularly spaced spherical bubbles of the same size are obtained. The upper surface is the broken (fragmented) surface. Even in this idealized case, each bubble (even if it is completely inside the magma) is not surrounded by a perfectly spherical shell. Instead, each bubble is surrounded by a melt shell with a spherical inner surface and polyhedral outer surface. **b**, Further approximation leads to spherical bubbles each surrounded by a spherical shell.

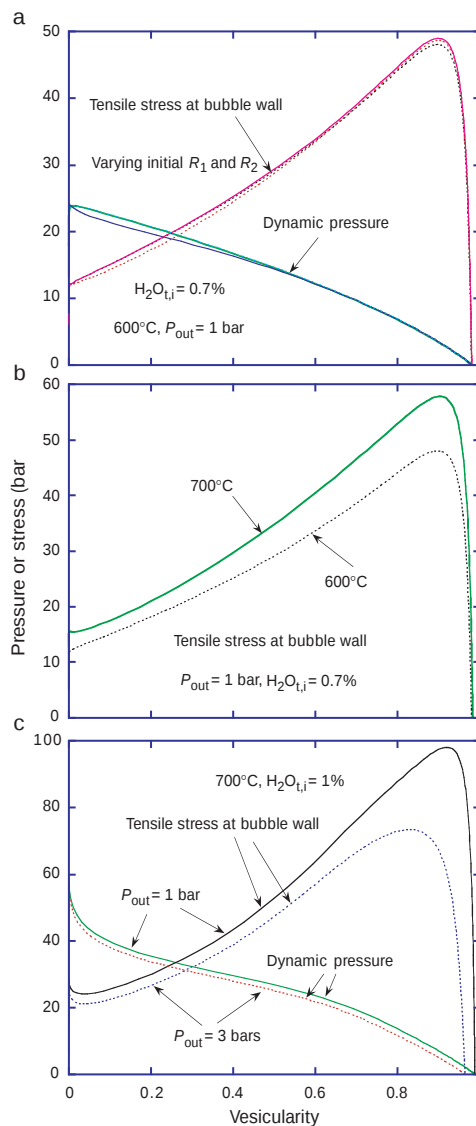


Figure 2 Calculated tangential stress (σ^{th}) in bubble walls as a function of vesicularity. **a,b,c**, Different experimental conditions. The program of Proussevitch and Sahagian¹² is used for calculation with H₂O diffusivity and solubility from Zhang¹⁷. The viscosity is that of Hess and Dingwell¹⁸ divided by two. The calculated stress may exceed the tensile strength because fragmentation is not incorporated into the calculation. The fragmentation point is obtained by comparing the calculated stress in the figure with the tensile strength. In **a**, the initial bubble radius R_1 varies from 1 to 5 μm and initial bubble+shell thickness R_2 varies from 50 to 200 μm .

without considering fractures. The calculated stress is then compared with tensile strength to determine the fragmentation point. Figure 2a shows that the calculated stresses are almost independent of the choice of the initial bubble radius (R_0) or initial melt volume surrounding each bubble (related to number density of bubbles). Hence the effect of the initial bubble size or number density is ignored. Figure 2b shows that the σ^{tt} versus x relation depends on the temperature. In a slowly growing dome when temperature decreases with time from 700 to 600 °C, it is expected that the σ^{tt} versus x relation would lie between the two trends in Fig. 2b. Figure 2c shows that σ^{tt} decreases as P_{out} increases. A comparison of Fig. 2b and c shows that the calculated stress increases with initial total H_2O ($\text{H}_2\text{O}_{\text{t,i}}$).

From Fig. 2, although P_{dyn} decreases with increasing vesicularity, σ^{tt} increases with x to roughly 90% vesicularity. The increase of σ^{tt} despite a decrease in ΔP is owing to the thinning of bubble walls (or decrease in x) so that ΔP must be accommodated in a small distance. Therefore, with gradual bubble growth in a slowly growing lava dome or a slowly advancing lava flow, the vesicularity may reach a critical value so that $\sigma^{\text{tt}} > S$, leading to a sudden fragmentation and crumbling of the dome or flow to a pyroclastic flow. The fragmentation conditions, depending strongly on $\text{H}_2\text{O}_{\text{t,i}}$, are listed in Table 1. The calculated vesicularity and final total H_2O level at fragmentation are reasonable values in pumice, suggesting that this criterion captures the essence of fragmentation. Furthermore, as long as $\text{H}_2\text{O}_{\text{t,i}}$ is high and the lava does not cool down rapidly, a dome or lava flow would eventually fragment when the stress reaches 50–60 bar. Hence, the composition $\text{H}_2\text{O}_{\text{t,i}}$ and temperature of a lava dome or flow can be used to forecast whether there will be fragmentation. The predicted vesicularity and final average total H_2O at fragmentation are negatively correlated when other conditions are equal but $\text{H}_2\text{O}_{\text{t,i}}$ is varied (Table 1). This prediction can be used to test my theory.

Factors that may help fragmentation of bubbly magma also include: a trigger to expose fresh interior bubbly lava owing to lava flow or gravitational instability may help fragmentation because of a sudden decrease in P_{out} ; or the falling of a piece of lava may send a strong enough shock wave through the piece to break bubble shells, leading to fragmentation.

The application to fragmentation in a volcanic conduit during an eruption is similar in principle, but in practice entails large uncertainties. One such uncertainty is the variation of P_{out} with time, which depends on the dynamics of an eruption, which in turn depends on bubble growth and nucleation rates. In the program¹², the decompression is linear, which is undoubtedly too simple^{19,20}. A second uncertainty is the number density of bubbles (that is, the choice of the initial shell thickness). Although it does not affect the stress versus vesicularity relation when P_{out} is constant (Fig. 2a), the number density (and hence new-bubble nucleation) plays an important role in modelling fragmentation in conduit eruptions. For example, for a given ascent rate, a greater number density implies more growth (and hence greater tensile stress) before the magma reaches an exit pressure of 1 bar.

In conclusion, on the basis of the maximum-stress theory for brittle failure, a bubbly magma fragments when the tensile stress at the inner walls of bubbles exceeds the tensile strength of the magma. This fragmentation criterion appears to effectively characterize the fragmentation of slowly growing lava domes and slowly advancing lava flows. Future work should include numerical-stress and bubble-growth calculations around bubbles with non-spherical shells, quantitative understanding of the strength of magma, and interaction between bubble growth/nucleation and conduit eruption dynamics.

Received 17 May; accepted 27 October 1999.

1. Sato, H., Fujii, T. & Nakada, S. Crumbling of dacite dome lava and generation of pyroclastic flows at Unzen volcano. *Nature* **360**, 664–666 (1992).
2. Fink, J. H. & Kieffer, S. W. Estimate of pyroclastic flow velocities resulting from explosive decompression of lava domes. *Nature* **363**, 612–615 (1993).

3. Navon, O., Chekhir, A. & Lyakhovsky, V. Bubble growth in highly viscous melts: theory, experiments, and autoexplosivity of dome lavas. *Earth Planet. Sci. Lett.* **160**, 763–776 (1998).
4. McBirney, A. R. & Murase, T. Factors governing the formation of pyroclastic rocks. *Bull. Volcanol.* **34**, 372–384 (1970).
5. Alidibirov, M. A. A model for viscous magma fragmentation during volcanic blasts. *Bull. Volcanol.* **56**, 459–465 (1994).
6. Proussevitch, A. A., Sahagian, D. L. & Kutolin, V. A. Stability of foams in silicate melts. *J. Volcanol. Geotherm. Res.* **59**, 161–178 (1993).
7. Papale, P. Strain-induced magma fragmentation in explosive eruptions. *Nature* **397**, 425–428 (1999).
8. Den Hartog, J. P. *Strength of Materials* (Dover, New York, 1977).
9. Lawn, B. *Fracture of Brittle Solids* (Cambridge University Press, 1993).
10. Zhang, Y. Mechanical and phase equilibria in inclusion-host systems. *Earth Planet. Sci. Lett.* **157**, 209–222 (1998).
11. Proussevitch, A. A., Sahagian, D. L. & Anderson, A. T. Dynamics of diffusive bubble growth in magmas: Isothermal case. *J. Geophys. Res.* **98**, 22283–22307 (1993).
12. Proussevitch, A. A. & Sahagian, D. L. Dynamics and energetics of bubble growth in magmas: Analytical formulation and numerical modeling. *J. Geophys. Res.* **103**, 18223–18251 (1998).
13. Webb, S. L. & Dingwell, D. B. Non-Newtonian rheology of igneous melts at high stresses and strain rates: experimental results for rhyolite, andesite, basalt, and nephelinite. *J. Geophys. Res.* **95**, 15695–15701 (1990).
14. Romano, C., Mungall, J. E., Sharp, T. & Dingwell, D. B. Tensile strengths of hydrous vesicular glasses: An experimental study. *Am. Mineral.* **81**, 1148–1154 (1996).
15. Alidibirov, M. & Dingwell, D. B. Magma fragmentation by rapid decomposition. *Nature* **380**, 146–148 (1996).
16. Liu, Y., Zhang, Y. & Sohn, J. Bubble growth in natural rhyolitic glasses. *Eos* **79**, F1013 (1998).
17. Zhang, Y. H_2O in rhyolitic glasses and melts: measurement, speciation, solubility, and diffusion. *Rev. Geophys.* **37**, 493–516 (1999).
18. Hess, K. & Dingwell, D. B. Viscosities of hydrous leucogranitic melts: A non-Arrhenian model. *Am. Mineral.* **81**, 1297–1300 (1996).
19. Zhang, Y., Sturtevant, B. & Stolper, E. M. Dynamics of gas-driven eruptions: Experimental simulations using CO_2 - H_2O -polymer system. *J. Geophys. Res.* **102**, 3077–3096 (1997).
20. Zhang, Y. Experimental simulations of gas-driven eruptions: kinetics of bubble growth and effect of geometry. *Bull. Volcanol.* **59**, 281–290 (1998).

Acknowledgements

This research is supported by the US NSF. I thank A. A. Proussevitch for making the bubble growth program available, J. R. Barber for discussion on brittle failure, Y. Liu for calculation of some bubble growth curves, and D. Snyder and P. Papale for comments.

Correspondence and requests for materials should be addressed to Y.Z. (e-mail: youxue@umich.edu).

Tube pumices as strain markers of the ductile–brittle transition during magma fragmentation

J. Marti*, C. Soriano* & D. B. Dingwell†

* Department of Natural Hazards, Institute of Earth Sciences, CSIC, Lluís Sole Sabaris s/n, 08028 Barcelona, Spain

† Bayerisches Geoinstitut, Universitaet Bayreuth, D-95440 Bayreuth, Germany

Magma fragmentation—the process by which relatively slow-moving magma transforms into a violent gas flow carrying fragments of magma—is the defining feature of explosive volcanism. Yet of all the processes involved in explosively erupting systems, fragmentation is possibly the least understood^{1,2}. Several theoretical and laboratory studies on magma degassing^{3–7} and fragmentation^{8–11} have produced a general picture of the sequence of events leading to the fragmentation of silicic magma^{12–14}. But there remains a debate² over whether magma fragmentation is a consequence of the textural evolution of magma to a foamed state where disintegration of walls separating bubbles becomes inevitable due to a foam-collapse criterion, or whether magma is fragmented purely by stresses that exceed its tensile strength. Here we show that tube pumice—where extreme bubble elongation is observed—is a well-preserved magmatic ‘strain marker’ of the stress state immediately before and during fragmentation. Structural elements in the pumice record the evolution of the magma’s mechanical response from viscous behaviour (foaming and foam

elongation) through the plastic or viscoelastic stage, and finally to brittle behaviour. These observations directly support the hypothesis that fragmentation occurs when magma undergoes a ductile–brittle transition and stresses exceed the magma's tensile strength.

In a mechanistic sense we may view the foam disaggregation scenario of magma fragmentation as a consequence of a cumulative strain of bubble growth. Thus, fragmentation should not be directly influenced by the rate of foaming but rather by its cumulative extent¹. For fragmentation of magma in this way a strain criterion must be overcome. Alternatively, magma may be fragmented by stresses that exceed its tensile strength. For melt-dominated systems, this implies that the strain rate in the magma during fragmentation approaches that which generates an unrelaxed or viscoelastic response of the melt. This mechanism thus contains a strain-rate criterion for fragmentation and is not dependent on the cumulative extent of vesiculation in order to meet it¹⁵. These two contrasting hypotheses have quite different implications for the conditions and consequences of the fragmentation process, as has been recently demonstrated in numerical simulations⁸.

An elegant way to evaluate fragmentation mechanisms, which is complementary to experimental and theoretical studies, is provided by the study of pumices^{1,16–18}. Explosive eruptions of foamed silicic magmas can produce two types of pumice clasts. These are the 'woody' pumice or 'tube' pumice, where extreme bubble elongation is preserved, and the 'spherical' pumice, where the bubbles have a more or less equant or spherical shape. Such studies must, however, make the assumption that the textures of the pumices faithfully record the physical state of the pyroclasts at fragmentation. Silicic magmas are highly viscous and the chemical and textural state of magma at fragmentation might well be preserved (effectively quenched) in certain cases. Textural analysis of pumice and ash fragments should, in such cases, indicate the physical conditions at fragmentation. Nevertheless, post-fragmentational chemical and textural evolution of spherical pumices has been demonstrated in some cases¹⁹. In contrast, analysis of tube pumices (see below) provides strong evidence that their texture is quenched at fragmentation. If so, tube pumice may well represent the best strain marker of eruptive processes at fragmentation that is available to us.

To evaluate this possibility, we provide here a combined chemical, physical and microstructural description of tube pumice from a deposit composed exclusively of crystal-poor tube pumices. We draw attention to microdomains where compressional strain is expressed as localized zones of ductile shear deformation before fragmentation.

The present samples were obtained from the Ramadas Caldera, a slightly elliptical (4×3 km), 8.7-Myr-old caldera (Central Andes, NW Argentina)^{20,21}, which generated a distinctive sequence of pumice fall deposits and minor ignimbrites. The pumice is entirely composed of lapilli-sized peraluminous rhyolitic tube pumices (with trace sub-millimetre pyroclasts) and lithic clasts of similar size. The Ramadas pumice clasts are highly angular, exhibiting little or no abrasion and rounding of the clast corners and no polishing of the tube pumice surfaces. Thus, there is no evidence of significant surface modification after fragmentation, and we may interpret preferred orientation in terms of the stresses operative during fragmentation.

Stretched bubbles are the most pervasive structure of the Ramadas pumice clasts. 40% of the tube pumices contain domains of shear deformation similar to those exhibited by multiply deformed foliated rocks (Fig. 1). Similar textures have previously been recorded (California)²², and we have also observed them elsewhere (Tenerife, Anatolia, Mexico). Domains of shear deformation consist of discrete dextral and sinistral shear planes which display a mean acute angle of 45° to the tube axes. Vesicle tubes are sheared along these planes which define a sort of cleavage crenulation (Fig. 1). No deformation of bubbles has been observed normal to the tube axes.

Deformation is always concentrated in those sections that are normal to the shear planes but contain the bubble stretching direction. Garnet crystals are mostly located on dextral shear planes. We suggest that they act as dislocation points for shear movements to be propagated on both types of planes. Also, many small circular vesicles occur around garnet crystals²³ that can be interpreted as strain shadows.

The Ramadas pumice clasts are typically terminated by surfaces that have one of two preferred orientations. The first, parallel to the tube axes, is typically adopted by two surfaces, at right angles. The second preferred orientation is at 45° – 60° angle to the tube axes, subparallel to the zones of localized shear deformation. The pumice clast surfaces may represent localization of the fragmentation along

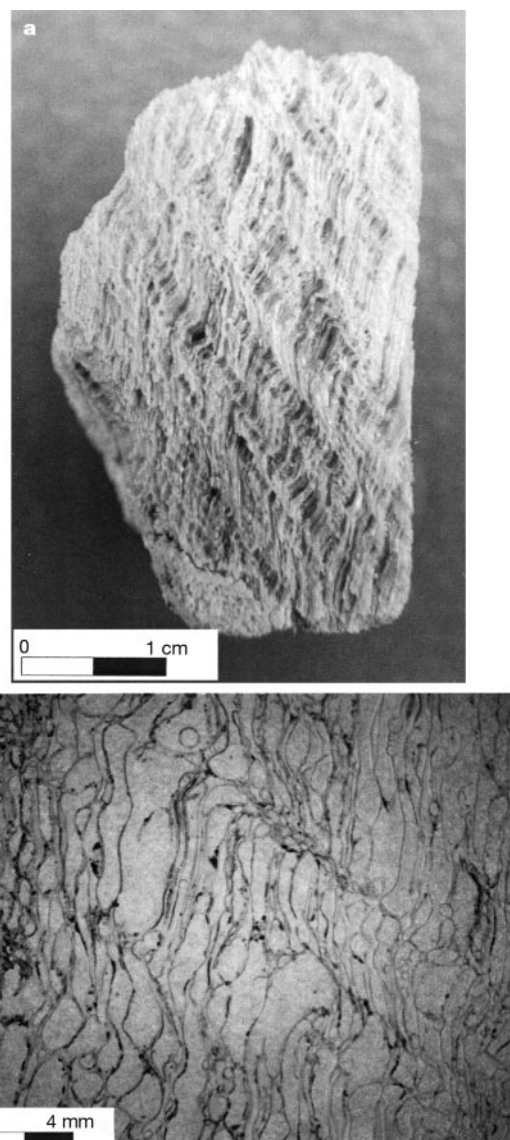


Figure 1 Texture of the tube pumice lapilli from Ramadas. **a**, Reflected-light and **b**, transmitted-light photomicrograph of polished thin section. Polished and thin sections were cut normal and parallel to the tubes and to the localized shear deformation which forms discrete planes at 45° to the tube axes. The tubes (average length 2 cm, width < 1 mm) are separated by thin crystal-free glass walls. Some bubbles are thinned and disrupted in a direction normal to the stretching direction. The elongated bubbles forming the tube pumice are deformed to create shear bands oriented at 45° to the bubble or tube elongation. Most fragment surfaces are aligned parallel to the 45° shear bands. The bubble elongation, shear bands and fragment surfaces are interpreted to indicate three successive stages of magma deformation: viscous, viscoelastic and brittle, resulting from increasing strain rates during magma ascent. The tube pumice acts as a strain marker across the fragmentation event in silicic explosive eruptions.

pre-existing zones of weakening due to localized shear deformation and resulting from prior longitudinal compression, or they may represent the brittle response to a tensile counterpart to the compressional features recorded by the shear bands. Clast morphology (sharp contacts and right angle intersections) and surfaces (cutting vesicle tubes) indicate that fragmentation occurred in the brittle regime of melt response. These stretched and sheared bubbles are strain markers of the instantaneous deformation immediately before fragmentation and rapid quenching. The strain axes can be assumed to be coaxial to the stress axes responsible for their formation and the strain markers can be analysed in terms of the instantaneous stresses during their formation. The texture and morphology of the Ramadas pumice clasts permits a strain analysis in two dimensions in planes containing shear deformation and the bubble-stretching direction (plane strain approximation, Fig. 2).

The presence of pyroclastic garnet (sp₂₄) indicates crystallization at 250–300 MPa and 860–875 °C (ref. 24). Further physical and chemical characteristics of the Ramadas tube pumice are density 0.70–0.88 g cm⁻³ (unsheared) versus 0.43–0.76 g cm⁻³ (sheared), corresponding to vesicularity ranges of 49–64 vol.% and 63–78 vol.% for the unsheared and sheared samples, respectively. Water contents are 2.3–4.9 wt %. X-ray diffraction analyses confirm the predominance of fresh pumice with little or no devitrification observed. Saturation of the Ramadas melt at the lower limit of these water contents requires a fragmentation pressure of 100 MPa for equilibrium degassing, and as low as 50 MPa for moderate levels of disequilibrium degassing⁴. The higher vesicularity range for the sheared pumices is consistent with the idea that the events generating shear bands occur at an advanced stage in the eruptive process and/or at advanced height in the conduit immediately before and below fragmentation.

The newtonian viscosity of the melt phase of the Ramadas tube pumice may be estimated as 10^{5.5} Pa s (refs 25, 26). The melt viscosity increases during the degassing process that generates the tubes but is approximately constant down to 2 wt% water. The bulk modulus of the foam can be estimated to be reduced by 4 log₁₀ units relative to the melt phase²⁷ to 10⁶ Pa, although the influence of the anisotropy on this estimate is unclear and the axial bulk modulus might be much closer to 10¹⁰ Pa.

The textures of the pumices (bubble shapes and sizes) suggest that bubble growth, stretching and shearing occur simultaneously. We propose that synchronous and coaxial stretching and shearing of bubbles can be explained by stresses parallel to the flow in the conduit. The growth and stretching of the bubbles would have occurred in the presence of an entirely viscous regime. The development of shear planes would also have occurred in the presence of viscous melt response. Bubble walls are continuous through the shear planes and there is no evidence for fragmentation and subsequent viscous healing. The stretching and growth of vesicles was further accompanied by both coalescence and disruption: further evidence of the ductile or viscous nature of pre-fragmentation deformation. Ductile coalescence is a process largely dominated by gas expansion in neighbouring vesicles, whereas ductile disruption is dominated by stretching of vesicles. Stretching of an initially small spherical vesicle and disruption of the stretched larger vesicle into smaller stretched ones represent a complete sequence of vesicle deformation during continuous expansion of bubbles in a ductile regime. Disruption has been observed not only in the stretching direction but also on both types of shear planes. Disruption is enhanced by strain rate and hindered by surface tension forces acting on the vesicles^{28,29}. Shearing of vesicles also occurs in the ductile regime, as demonstrated by the constant 45° angle between shear planes and the principal stress σ_1 (parallel to the tube axis) (Fig. 2) which is the angle experimentally predicted for shear failure in the ductile regime.

Vesicles smaller than 100 μ m have remained spherical, while larger vesicles have been stretched. This permits an estimation of

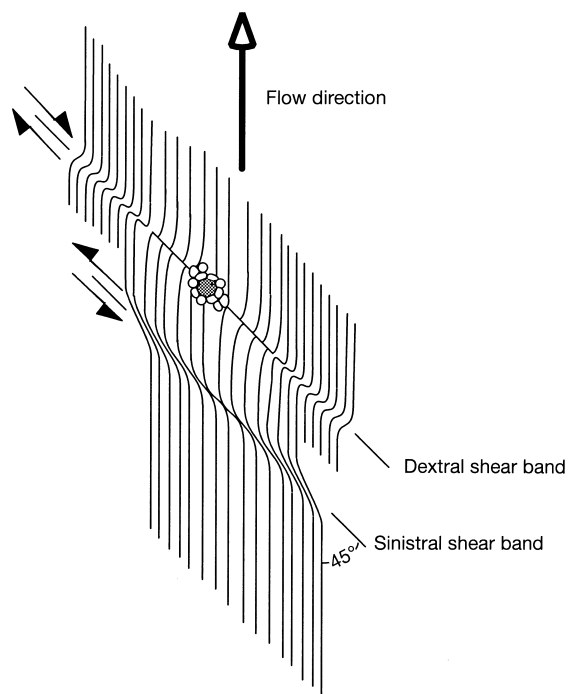


Figure 2 Synthetic pattern of the deformation structures observed in pumice clasts of the Ramadas pyroclastic deposit. Lines represent bubble walls. Shear bands (dextral and sinistral, indicated) develop at an angle of 45° to the stretching direction, which is parallel to the flow direction. The role of garnet crystals (hatched area) in shear band propagation and in the nucleation of small circular bubbles is also indicated.

the strain rate of bubble stretching using the critical radius criterion for bubble deformation^{26,28}. Applying the value of 10^{5.5} Pa s calculated above for the melt viscosity at eruptive temperatures, the strain rate inferred to preserve the spherical shape of bubbles at 100 μ m is 10^{-2.5} s⁻¹.

We interpret the localization of shear strain in the discrete planes to be the deformational response of a magma that exhibits a significant component of rigidity on the timescale of the shear deformation event. Therefore the shear deformation is interpreted as occurring in the domain of viscoelastic response of the magma, where rigidity is measurable but some viscous dissipation of the shear stresses occurs. The preservation of the very delicate and highly unstable geometry of the bubble walls in the shear planes is consistent with the proposal that the magma, following some dissipation of the shear stress, maintained its rigidity up to the point of fragmentation and thermal quenching to the glassy state. The deformation of the bubble walls in the shear planes may well have occurred under conditions where the viscosity of the melt phase was shear thinned during the deformation event, such that a non-newtonian melt rheology ensued^{2,8,15}. Both the stretching and the shearing of the bubbles preceded magma fragmentation, as both are cut by pumice edges. However, the total strain occurring between bubble shearing and fragmentation was clearly less than that required to recover the stable tubular geometry of the stretched bubbles in the vicinity of the shear planes. For a melt viscosity of 10^{5.5} Pa s, the strain rate necessary to induce viscoelastic or plastic response of the melt phase can be estimated from the parametrization of the onset of non-newtonian melt viscosity¹⁵ to be 10^{-2.5} s⁻¹ (the bulk modulus is 10⁶ Pa) to 10^{1.5} s⁻¹ (the bulk modulus is 10¹⁰ Pa). The lower strain rates estimated here are similar to the strain rate estimated for the tube flow. Thus the transition from ductile to brittle response may be viewed as a mechanical transition in the response of the tube pumice under a continuous stress regime during eruption.

Linkage of the shear features to flow and fragmentation dynamics is more difficult and will require an expanded experimental and simulation basis. Nevertheless, recent experimental studies of the fragmentation of tube pumices at high temperature and pressure indicate that stresses of the order of several megapascals are sufficient to overcome the strength of these volcanic materials³⁰. Such stresses should be easily achievable in explosive eruptions where the strain rate criterion is likely to be the deciding factor in the ease of fragmentation. Those experiments³⁰, in which fragmentation is demonstrably brittle, also generate tube pumice fragments with characteristic terminations at 90° and 45°–60° to the tube elongation. Shear bands are not seen in those single-stress-pulse experiments; this leads us to propose a multiple-stress-pulse origin for the shear bands observed here in Ramadas pumice.

Thus, the textures observed in the tube pumices from Ramadas are consistent with the proposal that the bubble stretching, bubble shearing and fragmentation of this pumice have occurred under strain rates that correspond to viscous, viscoelastic and brittle response regimes of the magma, respectively. Tube pumices are common products of explosive eruptions of silicic magmas, frequently appearing together with spherical pumices. Such shear bands have also been observed in phonolitic pumice indicating that eruptive conditions, not chemistry, control their generation. Further experimental and theoretical work is needed to quantify these aspects of magma fragmentation in order to understand one of the most enigmatic aspects of volcanic eruptions. □

Received 11 May; accepted 4 November 1999.

- Cashman, K. V. & Mangan, M. T. in *Volatiles in Magmas* (eds Carroll, M. R. & Holloway, J. R.) 449–478 (Mineral. Soc. Am. Rev. Mineral. 30, 1994).
- Dingwell, D. B. Volcanic dilemma: flow or blow? *Science* **273**, 1054–1055 (1996).
- Toramaru, A. Numerical study of nucleation and growth of bubbles in viscous magmas. *J. Geophys. Res.* **100**, 1913–1931 (1995).
- Proussevitch, A. A. & Sahagian, D. L. Dynamics and energetics of bubble growth in magmas: analytical formulation and numerical modelling. *J. Geophys. Res.* **103**, 18223–18251 (1998).
- Barclay, J., Riley, D. & Sparks, R. S. J. Analytical models for bubble growth during decompression of high viscosity magmas. *Bull. Volcanol.* **57**, 422–431 (1995).
- Bagdassarov, N., Dingwell, D. B. & Wilding, M. Rhyolite magma degassing: an experimental study of melt vesiculation. *Bull. Volcanol.* **57**, 587–601 (1996).
- Lyakhovsky, V., Hurwitz, S. & Navon, O. Bubble growth in rhyolitic melts: experimental and numerical investigation. *Bull. Volcanol.* **58**, 19–32 (1996).
- Papale, P. Strain-induced magma fragmentation in explosive eruptions. *Nature* **397**, 425–428 (1998).
- Mader, H. M., Phillips, J. C., Sparks, R. S. J. & Sturtevant, B. Dynamics of explosive degassing of magma: observations of fragmenting two-phase flows. *J. Geophys. Res.* **101**, 5547–5560 (1996).
- Alidibirov, M. & Dingwell, D. B. Magma fragmentation by rapid decompression. *Nature* **380**, 146–148 (1996).
- Klug, C. & Cashman, K. V. Permeability development in vesiculating magmas: implications for vesiculation, foam formation and fragmentation in lava fountains. *Bull. Volcanol.* **58**, 87–100 (1996).
- Dingwell, D. B. in *From Magma to Tephra: Modelling Physical Processes of Explosive Volcanic Eruptions* (eds Freundt, A. & Rosi, M.) 1–23 (Springer, Berlin, 1998).
- Dingwell, D. B. in *The Physics of Explosive Volcanic Eruptions* (eds Gilbert, J. & Sparks, R. S. J.) 1–26 (The Geological Society of London, Special Publication 145, 1998).
- Mader, H. M. in *The Physics of Explosive Volcanic Eruptions* (eds Gilbert, J. & Sparks, R. S. J.) 51–71 (The Geological Society of London, Special Publication 145, 1998).
- Dingwell, D. B. & Webb, S. L. Structural relaxation in silicate melts and non-Newtonian melt rheology in igneous processes. *Phys. Chem. Mineral.* **16**, 508–516 (1989).
- Heiken, G. & Wohletz, K. *Volcanic Ash* 246 (University of California Press, Berkeley, 1984).
- Houghton, B. F. & Wilson, C. J. N. A vesicularity index for pyroclastic deposits. *Bull. Volcanol.* **51**, 451–462 (1989).
- Dellino, P. & La Volpe, L. Fragmentation versus transportation mechanisms in the pyroclastic sequence of Monte Pilato-Rocche Rosse (Lipari, Italy). *J. Volcanol. Geotherm. Res.* **64**, 211–231 (1995).
- Thomas, N., Jaupart, C. & Vergnolle, S. On the vesicularity of pumice. *J. Geophys. Res.* **99**, 15633–15644 (1994).
- Viramonte, J. G. et al. Edad, genesis y mecanismos de erupción de las riolitas granatíferas de San Antonio de los Cobres, Provincia de Salta (Argentina). *Actas IX Congreso Geológico Argentino, Bariloche, Argentina* **3**, 216–233 (1984).
- Viramonte, J. G., Reynolds, J. H., Del Papa, C. & Disalvo, A. The Corte Blanco garnetiferous tuff: A distinctive late Miocene marker bed in northern Argentina applied to magnetic polarity stratigraph in the Rio Yacones, Salta province. *Earth Planet. Sci. Lett.* **121**, 519–531 (1994).
- Heiken, G. Plinian eruptions in the Medicine Lake Highland, California, and the nature of underlying magma. *J. Volcanol. Geotherm. Res.* **4**, 375–402 (1978).
- Heiken, G. & Eichelberger, J. R. Eruptions at Chaos Crags, Lassen Volcanic National Park, California. *J. Volcanol. Geotherm. Res.* **7**, 443–481 (1980).
- Gauthier, P. J., Déruelle, B., Viramonte, J. G. & Aparicio, A. Grenats des rhyolites de la caldera de La Pava-Ramada (NW Argentina) et leurs xénolites granitiques. *C. R. Acad. Sci.* **318**, 1629–1635 (1994).
- Hess, K.-U. & Dingwell, D. B. Viscosities of hydrous leucogranitic melts: a non-Arrhenian model. *Am. Mineral.* **81**, 1297–1300 (1996).

- Dingwell, D. B., Bagdassarov, N. S., Bussod, G. Y. & Webb, S. in *Experiments at High Pressure and Applications to the Earth's Mantle* (ed. Luth, R. W.) 131–196 (Mineral. Assoc. Can. Short Course Handbook 21, 1993).
- Chouet, B. A. *Monitoring and Mitigation of Volcano Hazards* (eds Scarpa, R. & Tilling, R.) (Springer, Berlin, 1996).
- Stein, D. J. & Spera, F. J. Rheology and microstructure of magnetic emulsions: theory and experiments. *J. Volcanol. Geotherm. Res.* **49**, 157–174 (1992).
- Dingwell, D. B. in *Structure and Dynamics of Silicate Melts* (eds Stebbins, J. E., McMillan, P. W. & Dingwell, D. B.) 21–66 (Mineral. Soc. Am. Rev. Mineral. 32, (1995).
- Martel, C., Dingwell, D. B., Spieler, O., Pichavant, M. & Holtz, F. Systematic foaming and fragmentation in synthetic melts. *Eos* **79**, F981 (1998).

Acknowledgements

We thank J. G. Viramonte for discussions on the Ramadas eruption and field assistance. This work was supported by European Union, DAAD, DFG (Di 431) and MEC grants.

Correspondence and requests for materials should be addressed to D.B.D. (e-mail: don.dingwell@uni-bayreuth.de).

Large-amplitude cycles of *Daphnia* and its algal prey in enriched environments

Edward McCauley*, Roger M. Nisbet†, William W. Murdoch‡, Andre M. de Roos§ & William S. C. Gurney||

* Ecology Division, Department of Biological Sciences, University of Calgary, Calgary, Alberta, Canada T2N 1N4

† Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, USA 93106

‡ Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, USA 93106

§ Section Population Biology, University of Amsterdam, Kruislaan 320, 1098 SM Amsterdam, The Netherlands

|| Department of Statistics and Modelling Sciences, University of Strathclyde, Glasgow, Scotland, UK

Ecological theory predicts that stable populations should yield to large-amplitude cycles in richer environments^{1–3}. This does not occur in nature. The zooplankton *Daphnia* and its algal prey in lakes throughout the world illustrate the problem^{4–6}. Experiments show that this system fits the theory's assumptions^{7–9}, yet it is not destabilized by enrichment⁶. We have tested and rejected four of five proposed explanations¹⁰. Here, we investigate the fifth mechanism: inedible algae in nutrient-rich lakes suppress cycles by reducing nutrients available to edible algae. We found three novel results in nutrient-rich microcosms from which inedible algae were excluded. First, as predicted by theory, some *Daphnia*-edible algal systems now display large-amplitude predator-prey cycles. Second, in the same environment, other populations are stable, showing only small-amplitude demographic cycles. Stability is induced when *Daphnia* diverts energy from the immediate production of young. Third, the system exhibits coexisting attractors—a stable equilibrium and large-amplitude cycle. We describe a mechanism that flips the system between these two states.

Analyses of many natural populations show that *Daphnia* strongly suppresses its edible algal food supply and thus *Daphnia* density is, in turn, limited as a result of reduced fecundity^{4,5,10}. We compared¹⁰ observed edible algal equilibrium densities, in lakes covering a wide range of nutrient levels, with algal densities predicted by a well parameterized model based on *Daphnia* energetics. We rejected four explanations for the absence of large-amplitude predator-prey cycles (predator-dependent functional

response¹¹, inedible algal interference⁵, density-dependent¹² or productivity-dependent herbivore mortality^{5,13}). We suggested that, instead, inedible algae compete successfully with edible algae in richer environments, thereby reducing the effective carrying capacity of *Daphnia*'s prey in productive environments. This would explain the mismatch between theory^{14–16} and empirical results⁶. Similar mechanisms involving edibility of prey have been proposed for other communities¹⁷. It was impossible to test this idea

using equilibrium relationships because inedible algal biomass in an enriched system¹⁸ increases with productivity, and therefore obscures its effect. We therefore examined *Daphnia*-edible algal dynamics in nutrient-rich environments from which inedible algae were completely excluded.

We achieved this manipulation in a series of replicate microcosms containing diverse plankton communities (bacteria, edible algal species, *Daphnia pulex*) minus inedible algae. In enriched systems containing inedible algae, our observations from lakes^{4,6}, mesocosm experiments⁶, and two sets of new experiments (data not shown; E.M. *et al.*, manuscript in preparation), have shown that *Daphnia*-edible algal dynamics do not display large-amplitude cycles. These systems are stable and can display stage-structured small-amplitude cycles. In contrast, results from our nutrient-rich microcosms without inedible algae reveal new features of the dynamics. First, as predicted by theory, *Daphnia* and their algal prey can indeed display large-amplitude predator–prey cycles at high nutrient levels and in the absence of inedible algae (Fig. 1). Fluctuations in *Daphnia* biomass exceeded a factor of five between peaks and troughs, with densities of more than 400 individuals per litre at the zenith of the cycles. Four systems in total exhibited this type of cycle. These large-amplitude cycles have never been observed before in the *Daphnia*–algal system. Second, under the same global environmental conditions (Fig. 1), some (6) of the populations display small-amplitude cycles (maximum/minimum *Daphnia* biomass less than 2). The difference in dynamics is not merely quantitative (that is, differing amplitudes); the two types are also characterized by differences in demography (Fig. 2). In the large-amplitude predator–prey (PP) cycles, juvenile and adult *Daphnia* fluctuate in synchrony ($P < 0.05$ cross-correlation time-series analysis) and suppression of fecundity results from increases in both juvenile and adult densities. By contrast, the small-amplitude cycles are stage-structured (SS) with a negative covariance between juveniles and adults, and fecundity suppression occurs as a result of the over-production of juveniles⁴. Thus large changes in amplitude and their demographic structure are distinguishing characteristics of the cycles.

This system appears to have two coexisting attractors, one associated with SS cycles and the other with PP cycles. This conclusion is reinforced by additional results from the first experiment: more than half of the systems (12 of 22) showed the two classes of dynamics in sequence (Fig. 3). Of these, eight of the

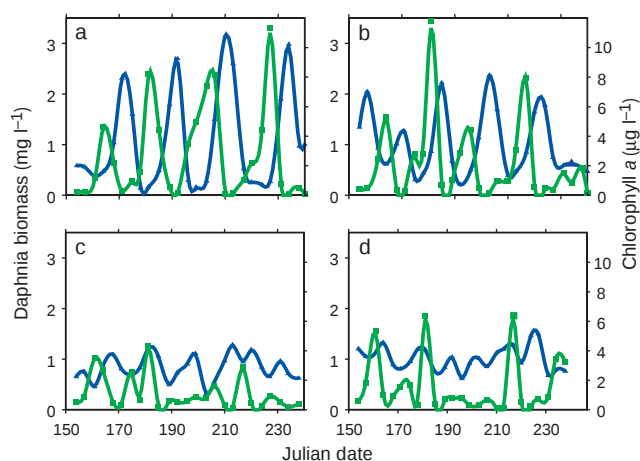


Figure 1 Large- and small-amplitude cycles in the same global environment. Population dynamics of *Daphnia* (blue triangles) and their edible algal prey (green squares) in four nutrient-rich systems from one treatment. The solid lines are spline fits to the time series (d.f., degrees of freedoms 20). **a, b**, Examples of large-amplitude predator–prey cycles. **c, d**, Examples of small-amplitude stage-structured cycles. The initial biomass of the replicates is similar. *Daphnia* biomass is calculated from estimates of density and size-structure, using length–weight relationships measured for the clone used in the experiment. Algal biomass is measured as chlorophyll *a* concentration.

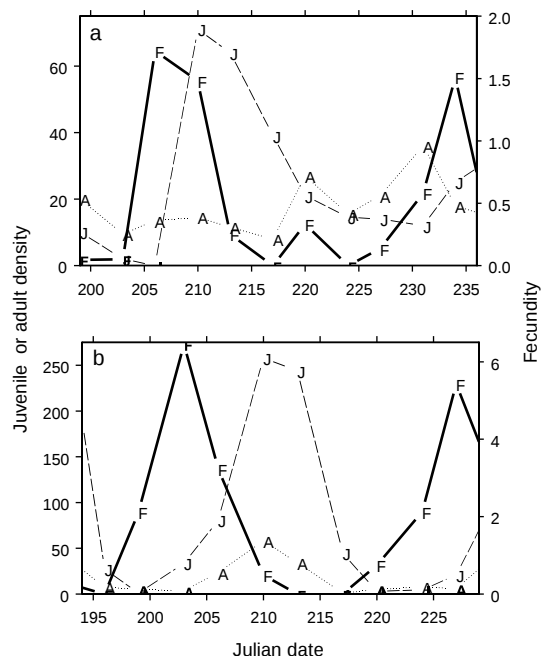


Figure 2 Suppression of *Daphnia*'s fecundity in predator–prey cycles and stage-structured cycles. In stage-structured cycles (**a**), *Daphnia*'s fecundity (F = eggs per adult) is suppressed primarily by the over-production of juveniles (individuals < 1.4 mm in length), whereas both juveniles (J) and adults (A) effect suppression in predator–prey cycles (**b**). Juveniles and adults co-vary positively during the predator–prey cycles, but not in the stage-structured cycles.

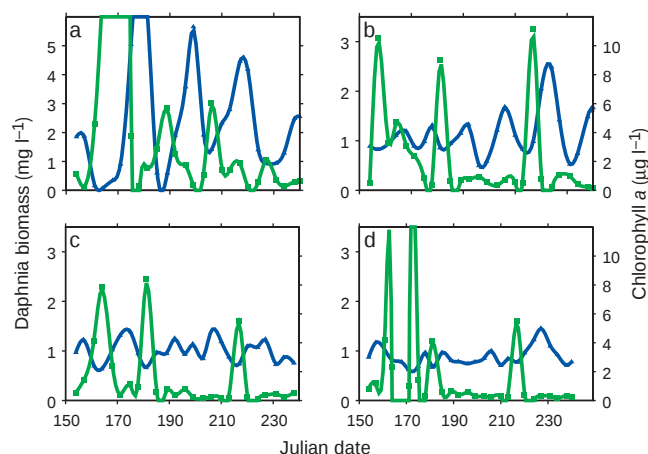


Figure 3 Systems can switch between large-amplitude and small-amplitude cycles. Population dynamics of *Daphnia* (blue triangles) and their edible algal prey (green squares) in four nutrient-rich systems from one treatment. The solid lines are spline fits to the time series (20 d.f.). **a, b**, Examples of systems displaying transitions from large-amplitude predator–prey cycle to stage-structured cycle (**a**) and small-amplitude cycle to large-amplitude predator–prey cycle (**b**). **c, d**, Systems with similar initial conditions from the same treatment that do not switch dynamics.

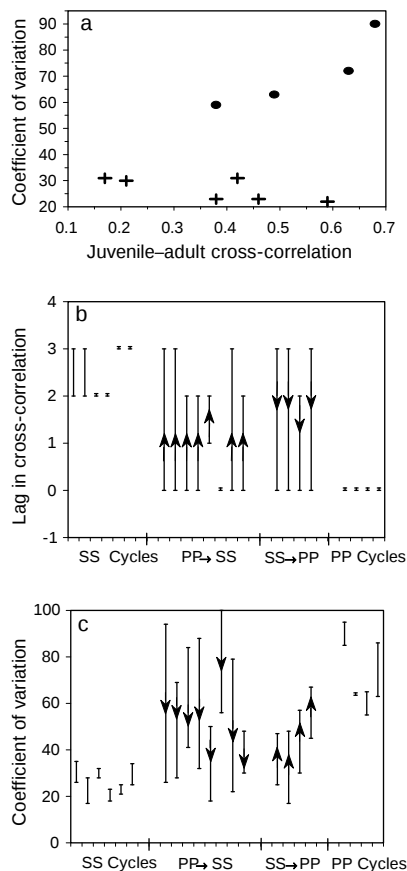


Figure 4 Detecting the different types of cycle. **a**, Shown is the difference between stage-structured cycles (SS) and predator–prey cycles (PP). SS cycles (crosses) have both a low coefficient of variation (reflecting a small amplitude) and peak values of cross-correlation coefficients between juveniles and adults occurring at a phase shift of three sampling periods (10–11 day temporal lag; see Fig. 2a). PP cycles (solid circles) have large amplitude, and juveniles and adults co-vary positively at zero lag (the peak value of the cross-correlation coefficient is at zero lag; see Fig. 2b). **b**, **c**, Shown is a comparison of these features for the first half of each series with the second half of the series. The horizontal bars show the starting and ending values; the lines and arrows show the magnitude and direction of the change that occurs between the first and second half of each series.

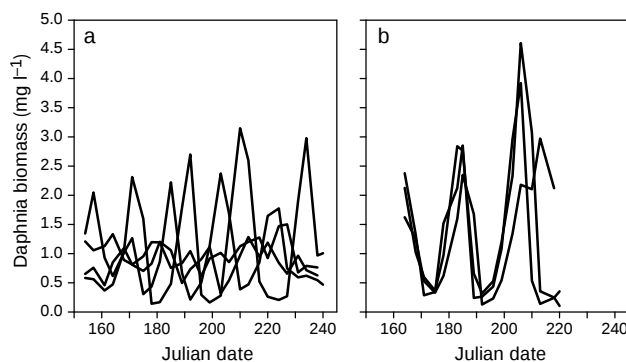


Figure 5 Energy channelling towards asexual reproduction destabilizes dynamics. **a**, Dynamics of replicate *Daphnia* populations illustrating that the replicates are asynchronous in time. **b**, Dynamics of the *Daphnia* populations in which energy channelling towards asexual reproduction is maintained by swapping adult females producing ephippial resting eggs with similarly sized females carrying asexual eggs.

systems showed PP cycles giving way to SS cycles, and four showed SS cycles giving rise to PP cycles (Fig. 4), based on comparing changes in both temporal variation and demographic correlations between juveniles and adults for the first and second half of each *Daphnia* series. As far as we know, this is the first experimental demonstration of coexisting attractors.

Why do only four of these populations show continuing large-amplitude cycles? Unless these systems are different from those analysed previously^{4,6,10}, the cause is unlikely to be direct density-dependence of *Daphnia*'s vital rates. The alternative is some mechanism that temporarily decouples *Daphnia*'s dynamics from the algal dynamics, and hence stabilizes the interaction, when food is scarce. One possibility is that in the more stable systems *Daphnia*'s rate of increase was reduced by the production of energy-intensive resting eggs (ephippia) that drop to the bottom and do not contribute to population growth. A variable fraction of the *Daphnia* population may produce ephippia in response to declining food supply, so ephippial production could prevent subsequent over-exploitation of the algal population and hence damp the large-amplitude cycle. It is well known that *Daphnia* can display alternate modes of reproduction in response to different environmental stimuli¹⁹ such as declining food per head, but the dynamical consequences of this feature have never been examined experimentally. We established that this mechanism operates in the following experiment.

We experimentally removed ephippia-producing females from one set of populations, replacing them with the same number of asexually-reproducing gravid females. Figure 5a shows the dynamics of replicate *Daphnia* populations where the energy channelling was not manipulated; the waxing/waning of the populations is not coincidental, which eliminates the possibility that common environmental variability drives the cycles. Replicate *Daphnia* populations where energy channelling was artificially directed to asexual reproduction (that is, away from resting egg production) displayed even larger-amplitude predator–prey cycles (Fig. 5b) with no tendency to show small-amplitude fluctuations or demography characteristic of the stage-structure cycles. Thus, ephippial production is one mechanism that can suppress predator–prey cycles at high nutrient levels in the absence of inedible algae.

We draw three conclusions with broad implications. First, given an environment that matched the theory, the *Daphnia*–algal system displayed large amplitude predator–prey cycles. We conclude that simple consumer–resource theory, that predicts such cycles in productive environments, is the appropriate starting point for investigations of population dynamics. Second, inedible algae may well suppress these inherent predator–prey cycles in natural waters. Competitors or other drains on resources may also explain stable consumer–resource interactions in other natural communities. Finally, ephippial production, which is stabilizing, is only one example of a stabilizing life-history feature that uncouples consumer dynamics from a temporarily scarce food supply. Other organisms respond to low resource density by shutting off reproduction and surviving better, which has a similar effect²⁰. The identification of uncoupling mechanisms that cause natural populations to maintain stability in enriched environments is a significant challenge for population ecologists. □

Methods

Dynamics of *Daphnia pulex* and edible algae were studied in 22 replicate systems under a 12 h light/12 h dark cycle at $22.0 \pm 0.3^\circ\text{C}$. Additionally, three control tanks were used to estimate the edible algal carrying capacity in the absence of *Daphnia*. Each 20 litre tank was filled with ultra-violet-sterilized water from Glenmore Reservoir, Alberta (pH 8.5). An edible algal 'cocktail' was created from stock cultures: *Selenastrum capricornum* (University of Toronto Culture Collection number 37), *Scenedesmus obliquus* (UTCC no. 5), *Mallomonas papillosa* (UTCCMP), *Ankistrodesmus convolutus* (UTCC no. 304), *Rhodomonas minuta* (UTCC no. 34), *Chlamydomonas reinhardtii* (University of Texas Culture Collection number 90), *Cosmarium* sp. (UTEX), *Cryptomonas erosa* (Freshwater Institute, Winnipeg) and added to each tank. Nutrient additions (KH_2PO_4 , NaNO_3 , $\text{Ca}(\text{NO}_3)_2 \cdot 4\text{H}_2\text{O}$) were based on empirical relationships¹⁸ to yield an edible algal carrying

capacity exceeding 3.0 mg carbon per litre (see below). After 7 days, the water from all tanks was mixed thoroughly to standardize algal initial conditions. Over the following 7 days, a fixed number of individually sized female *Daphnia pulex* drawn from a single-clone stock population was added to each tank to make starting conditions as uniform as possible.

Each tank was sampled twice weekly. Two replicate 1-litre water samples were drawn; gently filtered onto a 35- μ m mesh that was kept moist in a Petri dish. All *Daphnia* individuals in the sample were enumerated and sized, and then gently washed back into their tank. Sampling with replacement ensures that *Daphnia* death rates *per capita* were not shifted in parameter space. *Daphnia* biomass was calculated using density estimates for each size class, along with dry weight of individuals in each size class from length-weight relationships measured for the clone used in the experiment. Algal samples were filtered onto Whatman glass microfibre filters (GF/C), and frozen for later analysis following standard acetone extraction and fluorometric determination of chlorophyll *a* concentrations¹⁸. Samples preserved with Lugol's solution¹⁸ were inspected using an inverted-phase microscope to ensure that no inedible algal species invaded the tanks. Total phosphorus and soluble reactive phosphorus levels were routinely monitored along with biweekly measurements of pH and water hardness. Phytoplankton in replicate tanks with no *Daphnia* added did not display cyclic fluctuations.

To ensure no alternative food supply (such as algal or bacterial wall growth) was available for *Daphnia* in these nutrient-rich environments, all surfaces of the tanks were scraped daily for the entire duration of the experiments, and detrital material was vacuumed off the bottom every three days and treated to break down the material. The sample was ground using a high-speed micro-tissue grinder (Ultra Turrax Tissue Grinder, Staufen, Germany), sonicated for 20 min, and microwaved at high power for 3–5 min. The water sample was then cooled to the tank temperature and added back to the original tank. These procedures (scraping and regenerating bottom material) were effective in eliminating any wall growth, and any potential sink of organic material on the bottom of each tank.

Both simple biomass models^{8,21} and stage-structured predator–prey models⁷ parameterized using independent physiological data for *Daphnia* predict that large-amplitude prey–escape cycles should occur once the edible algal carrying capacity exceeds levels of approximately 0.2 mg carbon l⁻¹. Levels of nutrient enrichment were chosen¹⁸ to yield edible algal carrying capacities of more than 15 times that level (3.0 mg carbon l⁻¹), thus ensuring that the systems were placed deep in the parameter space that predicts predator–prey limit cycles.

To assess changes in dynamics between the two types of cycles (predator–prey and stage-structured) we compared two statistics—one describing the amplitude and the other describing the covariance of juveniles and adults. The coefficient of variation in *Daphnia* biomass (standard deviation/mean biomass) provides an indication of the relative magnitude of temporal variability as influenced by cycle amplitude, and the magnitude of cross-correlation coefficients between juvenile and adult biomass, estimated while one of the series is shifted (or lagged) in time, can be used to characterize cycle demography. As the demography of stage-structured cycles suggests (Fig. 2a), the cross-correlation between juveniles and adults is weak (in most cases slightly negative) for unlagged series and maximum when the juvenile series is shifted forward in time by approximately three sampling intervals (10–11 days). In predator–prey cycles (Fig. 2b), juveniles and adults co-vary positively, so that the maximum positive correlation occurs without any shifts in the series (that is, with zero lag). Thus, the stage-structured cycles have small coefficients of variation and maximum cross-correlations with a shift of 10–11 days (a lag of three sampling intervals), while the predator–prey cycles have large coefficients of variation and maximum cross-correlation coefficients at zero lag (Fig. 4a). Lags and coefficients of variation for particular systems can be assessed by comparing data from Fig. 4, b and c.

To test whether the production of ephippial resting eggs reduces the amplitude of predator–prey cycles, we set up six additional 20-litre tanks matched in pairs. Inoculations of tanks followed the protocol outlined above, except that the inoculation date for one tank in each pair was staggered by 2 weeks to ensure that the dynamics were out of phase. The exact cues for ephippial egg production are unclear, but appear to be related to rapidly declining food availability. Staggering tanks by 2 weeks helps to create donor tanks that contain asexual ovigerous females. We manipulated the realized energy channelling to asexual reproduction by swapping individual *Daphnia* that had produced ephippial eggs with ovigerous females from the donor tank of the pair. Tanks were checked every day and when a female carrying a resting egg was found, she was removed from the tank and replaced with a female of similar size carrying asexual non-resting eggs. Replacement with similarly sized individuals ensures that the numerical response is maintained without modifying the population-level ingestion rates. The tanks were sampled as described above. Experiments lasted until the supply of egg-bearing females in donor tanks was exhausted.

Received 19 May; accepted 9 September 1999.

- Rosenzweig, M. L. Paradox of enrichment: destabilization of exploitation ecosystems in ecological time. *Science* **171**, 385–387 (1971).
- May, R. M. Limit cycles in predator–prey communities. *Science* **177**, 900–902 (1972).
- Gurney, W. S. C. & Nisbet, R. M. *Ecological Dynamics* (Oxford Univ. Press, New York, 1998).
- McCauley, E. & Murdoch, W. W. Cyclic and stable populations: Plankton as paradigm. *Am. Nat.* **129**, 97–121 (1987).
- McCauley, E., Murdoch, W. W. & Watson, S. Simple models and variation in plankton densities among lakes. *Am. Nat.* **132**, 383–403 (1988).
- McCauley, E. & Murdoch, W. W. Predator–prey dynamics in environments rich and poor in nutrients. *Nature* **343**, 455–457 (1990).
- McCauley, E., Murdoch, W. W., Nisbet, R. M., de Roos, A. M. & Gurney, W. S. C. Structured population models of herbivorous zooplankton. *Ecol. Monogr.* **66**, 479–502 (1996).

- Nisbet, R. M. *et al.* in *The Art of Mathematical Modelling: Case Studies in Ecology, Physiology, and Biofluids* (eds Othmer, H. G., Adler, E. R., Lewis, M. A. & Dallon, J.) 61–79 (Prentice-Hall, Upper Saddle River, 1997).
- Andersen, T. *Pelagic Nutrient Cycles* (Springer, Berlin and Heidelberg, 1997).
- Murdoch, W. W., Nisbet, R. M., McCauley, E., de Roos, A. M. & Gurney, W. S. C. Plankton abundance and dynamics across nutrient levels: tests of hypotheses. *Ecology* **79**, 1339–1356 (1998).
- Arditi, R., Ginzburg, L. R. & Akçaya, H. R. Variation in plankton densities among lakes: a case for ratio-dependent models. *Am. Nat.* **138**, 1287–1296 (1991).
- Gatto, M. Some remarks on models of plankton densities in lakes. *Am. Nat.* **137**, 264–267 (1991).
- Persson, L. *et al.* Trophic interactions in temperate lake ecosystems: a test of food chain theory. *Am. Nat.* **140**, 59–84 (1992).
- Kretzschmar, M., Nisbet, R. M. & McCauley, E. A predator–prey model for zooplankton grazing on competing algal populations. *Theor. Popul. Biol.* **44**, 32–66 (1993).
- Grover, J. P. Competition, herbivory, and enrichment: nutrient-based models for edible and inedible plants. *Am. Nat.* **145**, 746–774 (1995).
- Abrams, P. A. & Walters, C. J. Invulnerable prey and the paradox of enrichment. *Ecology* **77**, 1125–1133 (1996).
- Bohannon, B. J. & Lenski, R. E. Effect of resource enrichment on a chemostat community of bacteria and bacteriophage. *Ecology* **78**, 2303–2315 (1997).
- Watson, S., McCauley, E. & Downing, J. A. Variation in algal community structure with enrichment. *Can. J. Fish. Aquat. Sci.* **49**, 2605–2610 (1992).
- Carvalho, G. R. & Hughes, R. N. The effect of food availability, female culture density, and photoperiod on ephippia production in *Daphnia magna* Straus. *Freshwat. Biol.* **13**, 37–46 (1983).
- Zonnefeld, C. & Kooijman, S. A. L. M. The application of a dynamic energy budget model to *Lymnaea stagnalis*. *Funct. Ecol.* **3**, 269–278 (1989).
- Nisbet, R. M., McCauley, E., de Roos, A. M., Murdoch, W. W. & Gurney, W. S. C. Population dynamics and element recycling in an aquatic plant–herbivore system. *Theor. Popul. Biol.* **40**, 125–147 (1991).

Acknowledgements

We thank E. Crone, L. Jackson, J. Post, W. Nelson and N. Hairston Jr for constructive criticism. M. Meding, J. LaMontagne and K. Fitzsimmons assisted with the experiment. S. Watson provided assistance with algal cultures and expertise on phytoplankton enumeration. E.M. acknowledges support from the Natural Sciences and Engineering Research Council of Canada.

Correspondence and requests for materials should be addressed to E.M. (e-mail: mccauley@uicgary.ca).

Ghrelin is a growth-hormone-releasing acylated peptide from stomach

Masayasu Kojima*, Hiroshi Hosoda*, Yukari Date*, Masamitsu Nakazato†, Hisayuki Matsuo* & Kenji Kangawa*

* Department of Biochemistry, National Cardiovascular Center Research Institute, Fujishirodai, Suita, Osaka 565-8565, Japan

† Third Department of Internal Medicine, Miyazaki Medical College, Miyazaki 889-1692, Japan

Small synthetic molecules called growth-hormone secretagogues (GHSs)^{1–3} stimulate the release of growth hormone (GH) from the pituitary^{4,5}. They act through GHS-R, a G-protein-coupled receptor for which the ligand is unknown. Recent cloning of GHS-R^{6,7} strongly suggests that an endogenous ligand for the receptor does exist and that there is a mechanism for regulating GH release that is distinct from its regulation by hypothalamic growth-hormone-releasing hormone (GHRH)^{4,5}. We now report the purification and identification in rat stomach of an endogenous ligand specific for GHS-R. The purified ligand is a peptide of 28 amino acids, in which the serine 3 residue is *n*-octanoylated. The acylated peptide specifically releases GH both *in vivo* and *in vitro*, and *O*-*n*-octanoylation at serine 3 is essential for the activity. We designate the GH-releasing peptide 'ghrelin' (*ghre* is the Proto-Indo-European root of the word 'grow'). Human ghrelin is homologous to rat ghrelin apart from two amino acids. The occurrence of ghrelin in both rat and human indicates that GH release from the pituitary may be regulated not only by hypothalamic GHRH, but also by ghrelin.

Cells expressing an orphan G-protein-coupled receptor (one with no known ligand) have been used as assay systems to identify unknown endogenous ligands^{8–12}. We constructed a stable CHO cell line expressing rat GHS-R (CHO-GHSR62) to monitor changes in the intracellular calcium concentration ($[Ca^{2+}]_i$) that were induced by rat tissue extracts from brain, lung, heart, kidney, stomach and intestine. As the highest activity was found in the stomach extract, this was further purified by successive chromatography^{13,14}, including Sephadex G-50 gel filtration (Fig. 1), CM-ion-exchange high-performance liquid chromatography (HPLC) and reverse-phase HPLC (RP-HPLC) (Fig. 2a, top panel). Gel filtration revealed potent $[Ca^{2+}]_i$ -increasing activity in the fractions of relative molecular mass (M_r) $\sim 3,000$ (3K; Fig. 1). Active fractions (fractions 43–48) were further purified by CM-ion-exchange HPLC, followed by RP-HPLC (Fig. 2a, top panel) to give 16 μ g of pure ghrelin from 40 g of rat stomach tissue.

The amino-acid sequence of the purified rat ghrelin was determined by a protein sequencer to be GSXFLSPEHQKAQQR-KESKKPPAKLQPR (X, not identified). Complementary DNA analysis of rat ghrelin indicated that the X residue at position 3 should be serine, while the rest of the sequence was identical to that determined by peptide sequencing. Peptide sequencing resulted in a very low yield of phenylthiohydantoin (PTH) at the step for X, suggesting that Ser 3 might be modified. However, a synthetic 28-residue peptide (X = Ser) having the sequence deduced from cDNA analysis did not increase $[Ca^{2+}]_i$ (Fig. 2b). Moreover, RP-HPLC showed that natural ghrelin eluted about 10 min later than the synthetic 28-residue peptide (desacyl ghrelin) (Fig. 2a, top and middle panels). This indicates that Ser 3 in natural ghrelin must be modified by a hydrophobic moiety.

To identify the hydrophobic moiety of Ser 3, we subjected purified ghrelin to electrospray ionization mass spectrometry. The observed M_r of purified ghrelin was $3,314.9 \pm 0.7$, 126 mass units higher than the M_r (3,189) calculated for the 28-residue sequence (X = Ser). This indicated that a hydrogen atom of the hydroxyl group in Ser 3 was replaced by a $C_7H_{15}CO$ moiety (127 mass). In other words, the hydroxyl group of Ser 3 residue is octanoylated. To verify the deduced structure, we synthesized an $[O-n\text{-octanoyl-Ser 3}]$ peptide, and found that it is identical to natural ghrelin in its behaviour under chromatography, mass-spectrometric pattern and GH-releasing activity. As clearly seen in Fig. 2a (bottom panel), natural ghrelin co-migrates with the synthetic n -octanoylated peptide on RP-HPLC. Furthermore, amino-terminal peptides (Gly-1~Phe-4) prepared by chymotrypsin digestion of both natural ghrelin and synthetic n -octanoylated peptide also showed the same retention times on RP-HPLC and identical fragmentation patterns when analysed on a time-of-flight mass spectrometer. In CHO-GHSR62 cells, synthetic ghrelin clearly induced an increase in $[Ca^{2+}]_i$ that was as potent as that induced by natural ghrelin (Fig. 2b), and stimulated GH secretion from rat primary pituitary

cells (described below). Thus, the complete structure of rat ghrelin is unambiguously identified as an $[O-n\text{-octanoyl-Ser 3}]$ -peptide (Fig. 2c). Octanoylation has not been observed previously in peptide modification. The fact that the n -octanoyl moiety is essential for the activity of ghrelin indicates that the post-translational mechanism for acyl modification should be urgently elucidated. It is also interesting that no structural homology with ghrelin is observed among synthetic GHSs.

Ghrelin potentially induced an increase in $[Ca^{2+}]_i$ in CHO-GHSR62 cells (50% excitatory concentration (EC_{50}) 2.5×10^{-9} M), whereas GHRH had no effect on the cells under the same conditions. The threshold for the ghrelin-induced increase in $[Ca^{2+}]_i$ was 10^{-11} M (Fig. 3), but the $[Ca^{2+}]_i$ increase was depressed in the presence of 10^{-4} M $[D\text{-Lys-3}]\text{-GHRP-6}$, an antagonist specific for GHS-Rs^{3,5}. A high dose of ghrelin, however, regenerated the $[Ca^{2+}]_i$

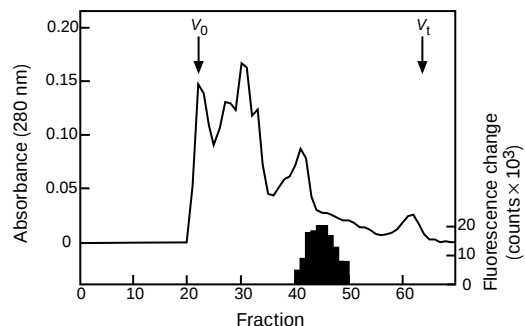


Figure 1 Gel filtration of Sephadex G-50 (fine) of SP-III fraction from 40 g rat stomach. Black bars indicate fluorescence change owing to $[Ca^{2+}]_i$ increase on CHO-GHSR62 cells. Active fractions were eluted around M_r 3,000. V_0 , void volume; V_t , total volume.

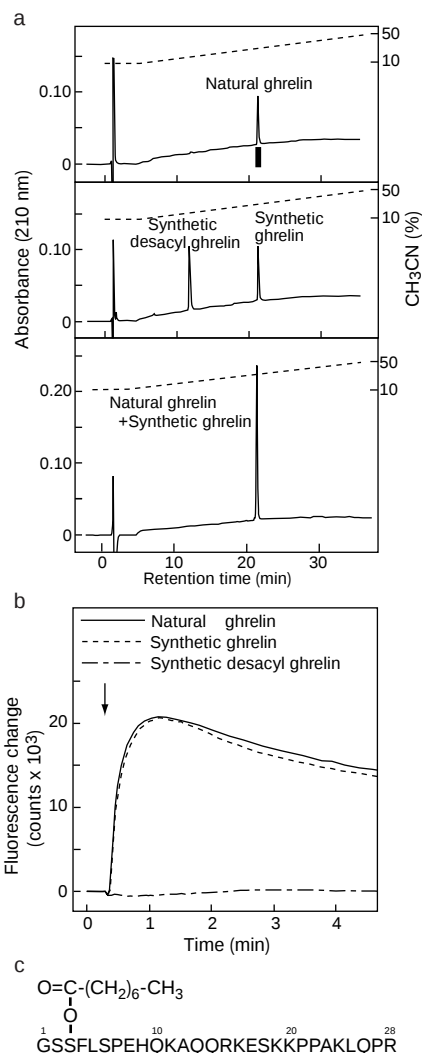


Figure 2 Identification of the n -octanoyl modification in ghrelin. **a**, Chromatographic comparison on RP-HPLC of natural ghrelin (top panel), synthetic ghrelin and desacyl ghrelin (middle panel), and co-migration of natural and synthetic ghrelin (bottom panel). Black bar indicates $[Ca^{2+}]_i$ -increasing activity (top panel). Desacyl ghrelin is a synthetic 28-residue peptide (X = Ser), which has the sequence deduced from cDNA analysis. Each peptide (2 μ g) was applied on a μ Bondasphere column (Waters). Flow rate, 1 ml min⁻¹. Solvent system, a linear gradient elution from A to B for 40 min. H₂O: CH₃CN: 10% TFA for A was 90:10:1, for B, 40:60:1 (v/v). Natural ghrelin co-migrates with synthetic ghrelin in the present conditions (bottom panel). **b**, Time courses for changes in $[Ca^{2+}]_i$ in CHO-GHSR62 cells induced by natural ghrelin, synthetic ghrelin and desacyl ghrelin. Each peptide (10^{-8} M) was added at the time indicated by the arrow. **c**, Structure of rat ghrelin.

change to the level observed in the absence of the antagonist (Fig. 3). These results indicate that ghrelin is competitively inhibited by the antagonist.

Although ghrelin had no sequence homology to any known biologically active peptides, we found an amino-acid sequence identical to that of rat ghrelin in a rat expressed sequence tag (EST; GenBank accession no. AI549172). On the basis of this sequence, we designed primers for the polymerase chain reaction (PCR) and amplified a cDNA fragment from rat stomach cDNA as the template. The fragment was used as a screening probe on a rat stomach cDNA library. A cDNA of 501 base pairs (bp) long was obtained. The predicted initiation methionine (nucleotides 31–33) conforms to Kozak's rules¹⁵. The open reading frame starting at this ATG codon encodes a 117-residue polypeptide, designated rat prepro-ghrelin (Fig. 4). The N-terminal 23-residue sequence of prepro-ghrelin exhibits features characteristic of a secretory signal peptide¹⁶. Ghrelin sequence starts from Gly24, which directly follows the signal peptide. The last two residues of ghrelin, Pro-Arg, should be a processing signal¹⁷.

Using rat ghrelin cDNA, we screened a human stomach cDNA library under low-stringency conditions¹⁸. Isolated human cDNA was found to encode human prepro-ghrelin of 117 amino acids (Fig. 4). Amino-acid identities between rat and human prepro-ghrelins are 82.9%. Only two amino acids are replaced in the 28-residue ghrelin segment, indicating that ghrelins are highly conserved between species. We purified human ghrelin from human stomach extract, and verified by synthesis that human ghrelin, like rat ghrelin has an *n*-octanoyl modification at Ser 3 (details will be reported

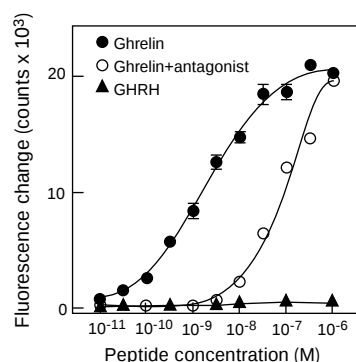


Figure 3 Dose–response relationships of $[Ca^{2+}]_i$ change in CHO-GHSR62 cells, in the presence or absence of $[D-Lys-3]$ -GHRP-6, a GHS-R-specific inhibitor. Inhibitor concentration was 10^{-4} M. GHRH was also assayed. Data points are means ($\pm$ s.e.m.) of triplicates for each experiment.

Human	1	MPSPGTVCSLLLLGMLWLDL	AMAGSSFLSP	30
Rat	1	MYSSATICSLLLLSMLWMDM	AMAGSSFLSP	30
Human	31	EHORVQORKESKPPAKLP	PRALAGWLRLPE	60
Rat	31	EHQKAQORKESKPPAKLP	PRALEGWLHPE	60
Human	61	DGGQAEAGAEDELEVRNAP	FDVGIKLSGVQ	90
Rat	61	DRGQAEAEAELEIRFNAP	FDVGIKLSGAQ	90
Human	91	YQOHSQALGKFLQDILWEE	AKEAPADK	117
Rat	91	YQOHGRLGKFLQDILWEE	VKEAPANK	117

Figure 4 Alignment of amino-acid sequences of human and rat prepro-ghrelin. Sets of identical residues are shaded. The dotted line indicates a signal peptide. The filled arrowhead indicates a cleavage site of a signal peptide. The open arrowhead indicates a C-terminal-processing site. Ghrelin sequences are boxed. Asterisks indicate *n*-octanoyl modification sites.

elsewhere).

An assay for GH-releasing activity using primary cultured pituitary cells^{4,5} revealed that ghrelin induced GH release in a clear dose-dependent manner (Fig. 5a, left panel). We also confirmed that GHRH showed GH-releasing activity in the present assay conditions (Fig. 5a, right panel). The GH-releasing potency of ghrelin is comparable to that of GHRH, EC_{50} values for GH-releasing activity being estimated as 2.1×10^{-9} M for ghrelin and 0.6×10^{-9} for GHRH. To confirm whether ghrelin specifically stimulates GH release, we analysed which hormones were released from pituitary cells by ghrelin. We found that the peptide specifically stimulated GH release, and did not affect other pituitary hormones, even at a high ghrelin dose (10^{-6} M) (Fig. 5b). Furthermore, when ghrelin was injected intravenously in anaesthetized rats, only GH concentration increased to the maximum levels at 5–10 min after injection (Fig. 5c). These results indicate that ghrelin is a GH-specific releasing peptide.

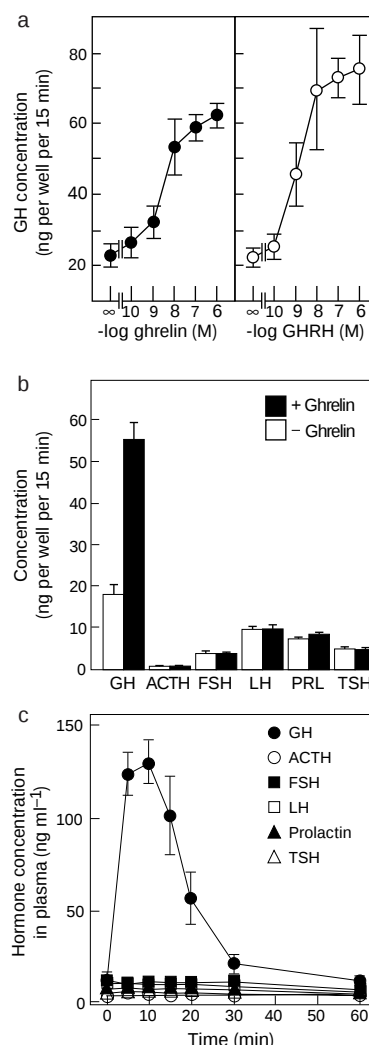


Figure 5 Effects of ghrelin on pituitary hormone secretion, *in vitro* and *in vivo*. **a**, Dose–response relationships in GH release from rat primary pituitary cells induced by ghrelin (left panel) and GHRH (right panel) (*in vitro*). GH concentration represents means ($\pm$ s.e.m.) of triplicate assays. **b**, Effects of a high-dose (10^{-6} M) ghrelin on hormone secretions from rat primary pituitary cells, *in vitro*. Each hormone concentration represents means ($\pm$ s.e.m.) of quadruplicate assays. **c**, Time courses of plasma hormone concentrations after intravenous injections of ghrelin into male rats. Synthetic ghrelin (10μ g) was injected intravenously in male rats (250 g) anaesthetized with pentobarbital. Blood samples were collected from cannulae in the femoral artery for up to 60 min and hormone concentrations were determined as described in Methods. Data represent means ($\pm$ s.e.m.) of four rats.

Northern blot analysis of rat tissue showed that prepro-ghrelin messenger RNA of 0.62 kilobases (kb) occurs in stomach (Fig. 6a). *In situ* hybridization indicated that ghrelin mRNA occurs in the region from the neck to the base of the oxyntic gland (Fig. 6b–d). Ghrelin-immunoreactive cells in the stomach have the same distribution as that found using *in situ* hybridization (Fig. 6e–g). The distribution pattern and morphological features of the labelled and

immunostained cells show that ghrelin cells are endocrine cells. Moreover, we found that ghrelin circulates in healthy human blood at a considerable plasma concentration (117.2 ± 37.2 fmol ml⁻¹; $n = 6$). Taken together with the fact that ghrelin, when injected intravenously, induces GH release (Fig. 5c), it is highly likely that this molecule is produced in and secreted from the stomach, circulating in the blood stream to act on the pituitary.

Although no detectable northern blot signal was observed in mRNA from whole brain (Fig. 6a), RT-PCR amplification of ghrelin mRNA revealed the presence of this transcript in brain. Immunohistochemical analyses performed after colchicine treatment¹⁹ revealed that ghrelin-immunoreactive neurons were localized in the hypothalamic arcuate nucleus (Fig. 6h, i). Considering that GHS-R mRNA is known to exist in hypothalamic regions, including the arcuate nucleus and pituitary gland²⁰, these results suggest that ghrelin in the arcuate nucleus may act on the hypothalamus or be transported to the anterior pituitary.

Thus, the occurrence of ghrelin in both stomach and hypothalamus will give a new dimension to the regulation of GH release. Further, our RT-PCR analysis indicates that GHS-R is also expressed in heart, lung, pancreas, intestine and adipose tissue (data not shown). Ghrelin may thus have multifaceted roles in, for example the cardiovascular system²¹ and metabolism²². □

Methods

Construction of GHS-R-expressing cell and assay system

Full-length cDNA of rat GHS-R was obtained by RT-PCR, with rat brain cDNA as the template. Sense and antisense primers were synthesized on the basis of the reported sequence of rat GHS-R⁷. The sense primer is 5'-ATGTGGAACGCGACCCCGACGGA-3', and the antisense primer is 5'-ACCCCAATTGTTTCCAGACCCAT-3'. The amplified cDNA was ligated into pcDNA3.1 vector (Invitrogen). The expression vector, GHSR-pcDNA3.1, was transfected into CHO cells, then stable expressing cells were selected using 1 mg ml⁻¹ G418. The selected cell line, CHO-GHSR62, responded to GHRP-6 (Peninsula Laboratories) at 10^{-10} – 10^{-9} M. Changes in intracellular Ca²⁺ concentration were measured using the FLIPR System (Molecular Device)^{23,24}. CHO-GHSR62 cells (4×10^4 cells) were plated into 96-well black-wall microplates (Corning) 12–15 h before assay. The cells were incubated with 4 μ M of fluorescent dye Fluo-4 AM (Molecular Probe) for 1 h, then washed 4 times with Hank's BSS containing 20 mM HEPES and 2.5 mM probenecid. Samples were tested to induce changes in fluorescence. Maximum [Ca²⁺]_i changes were determined in triplicate.

Purification of ghrelin

Fresh rat stomach (40 g) was diced and boiled for 5 min in 5 volumes of water to inactivate intrinsic proteases^{13,14}. The solution was adjusted to 1 M AcOH, 20 mM HCl. Peptides were extracted by homogenizing with a Polytron mixer. The supernatant of the extracts, obtained after 30 min centrifugation at 11,000 r.p.m., was concentrated to about 40 ml by an evaporator. The residual concentrate was subjected to acetone precipitation at a concentration of 66% acetone. After the precipitates had been removed, the supernatant acetone was evaporated. It was loaded onto a 10-g cartridge of Sep-Pak C18 (Waters), which was pre-equilibrated with 0.1% trifluoroacetic acid (TFA). The Sep-Pak cartridge was washed with 10% CH₃CN/0.1% TFA, and then eluted with 60% CH₃CN/0.1% TFA. The eluate was evaporated and lyophilized. The residual materials were redissolved in 1 M AcOH and then adsorbed on a column of SP-Sephadex C-25 (H⁺ form), pre-equilibrated with 1 M AcOH. Successive elutions with 1 M AcOH, 2 M pyridine and 2 M pyridine-AcOH (pH 5.0) provided three fractions of SP-I, SP-II and SP-III. Lyophilized SP-III fraction was applied on a Sephadex G-50 gel-filtration column. A portion of each fraction was subjected to the [Ca²⁺]_i-change assay using CHO-GHSR62 cells. Active fractions (43–48) were then separated by CM-ion-exchange HPLC on a column of TSK CM-2SW (4.6 $\times$ 250 mm; Tosoh) at pH 6.4. CM-HPLC active fractions were further fractionated by the second CM-HPLC on the same column at pH 4.8. The active peak was finally purified manually by reverse-phase HPLC using μ Bondasphere C 18 (3.9 $\times$ 150 mm; Waters) column. The amino-acid sequence of the peptide was analysed with a protein sequencer (494; Applied Biosystems).

Syntheses of rat and human ghrelins

The fully protected peptide, except for the hydroxyl group of Ser 3, was synthesized by the Fmoc solid-phase method on a peptide synthesizer (433A; Applied Biosystems). The hydroxyl group of Ser 3 in the peptide obtained above was then acylated with *n*-octanoic acid by the action of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide in the presence of 4-(dimethylamino) pyridine.

Cloning of rat and human ghrelin cDNAs

On the basis of the amino-acid sequence determined for purified ghrelin, we searched GenBank's EST database (dEST). One rat EST sequence (accession no. A1549172) contained the ghrelin sequence. On the basis of this EST sequence, we designed sense and

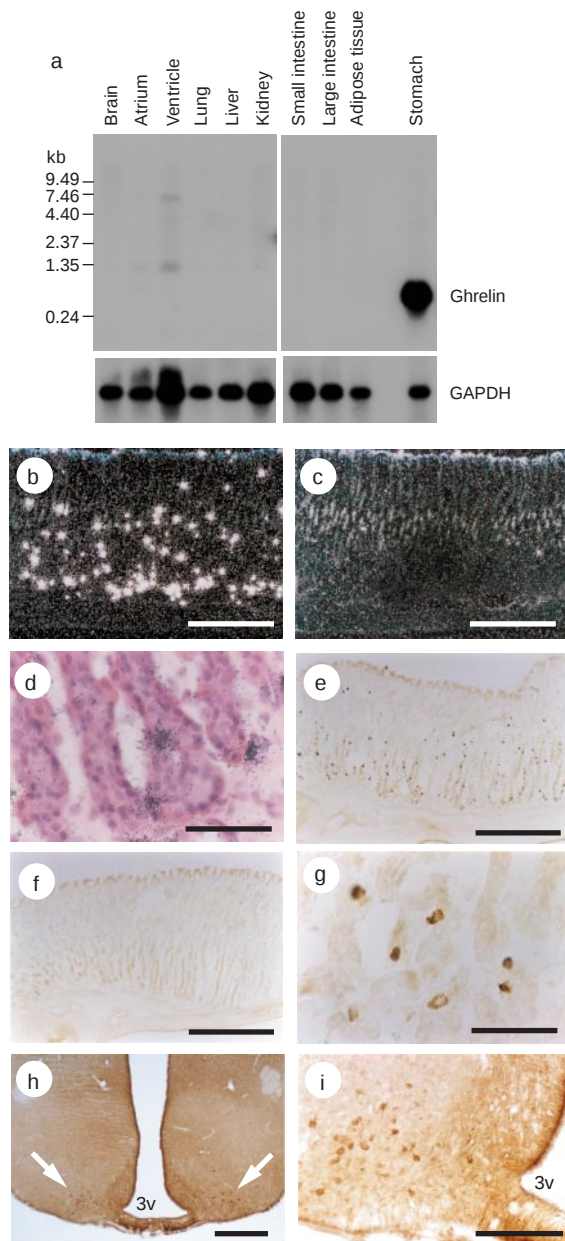


Figure 6 Expression studies of ghrelin. **a**, Northern blot analysis of ghrelin mRNA in rat tissue. Each lane contained 4 μ g of poly(A)⁺ RNA. ³²P-labelled rat ghrelin cDNA was used for hybridization. Autoradiography was exposed for 1 day. The lower panel indicates hybridization with glutaraldehyde-3-phosphate dehydrogenase (GAPDH) cDNA as internal control. **b–i**, Ghrelin expression in rat stomach (**b–g**) and brain (**h, i**). **b**, Hybridization signals in the oxyntic gland by the addition of excess of the unlabelled probes. **d**, High magnification of **b** (with haematoxylin counterstaining). **e**, Ghrelin-immunoreactive cells in the stomach have the same distribution as found on *in situ* hybridization. **f**, No immunoreactivity in the stomach after absorption of ghrelin antiserum. **g**, High magnification of **e**. **h**, Arrows indicate ghrelin neurons in the ventral–lateral border of the hypothalamic arcuate nucleus in colchicine-treated rats. 3v, third ventricle. **i**, High magnification of **h**. Scale bars: **b, c, e, f**, 400 μ m; **d, g**, 40 μ m; **h**, 200 μ m; **i**, 500 μ m.

antisense primers. Sense primer: 5'-TTGAGCCCAGAGCACCAGAAA-3'. Antisense primer: 5'-AGTTGCAGAGGAGGCAGAAGCT-3'. RT-PCR was performed with rat stomach cDNA as the template. The conditions for PCR were 35 cycles of 98 °C for 10 s, 55 °C for 30 s and 72 °C for 1 min. The amplified fragment was used as a screening probe for a rat stomach cDNA library. Approximately 2×10^5 recombinant phages were screened. Both strands of cloned cDNAs of rat ghrelin were sequenced. Human stomach cDNA library was constructed from 1 µg of human stomach poly(A)⁺RNA (Clontech) with a cDNA synthesis kit (Pharmacia). Full-length cDNAs of human ghrelin cDNA were isolated from the human stomach library, using rat ghrelin cDNA as the screening probe.

In vitro assay of pituitary hormone releasing activity

This was performed mainly by a previously described method^{4,5}. Anterior pituitaries were obtained from 4-week-old male Sprague-Dawley rats. After dispersion by collagenase, cells were collected and washed twice with Dulbecco's modified Eagle's medium (DMEM) containing 10% FCS and antibiotics²⁵. Cells were resuspended in an appropriate volume of the above medium, and 5×10^4 cells were seeded in poly-L-lysine-coated 96-well tissue-culture plates. The cells were cultured for 3–4 days. The culture medium was then replaced by 0.1 ml of DMEM containing samples, and the cells were incubated for 15 min at 37 °C. Aliquots of the resulting incubation medium were subjected to RIA or ELISA. Concentrations of pituitary hormones were measured with assay kits. Biotrak RIA (Amersham): GH, follicle-stimulating hormone (FSH), luteinizing hormone (LH), prolactin (PRL) and thyroid-stimulating hormone (TSH); high sensitivity EIA Kit (Peninsula Laboratories): adrenocorticotropin (ACTH).

Immunohistochemistry

Polyclonal antibody was raised in rabbit against the carboxy-terminal fragment (13–28) of ghrelin. [Cys⁰]-ghrelin (13–28) was coupled to maleimide-activated mariculture keyhole limpet haemocyanin (mCKLH; Pierce), and the conjugate was used for immunization. Immunohistochemical analyses were carried out as described elsewhere²⁶. Colchicine was injected into the rats' lateral ventricle through implanted cannulae 30 h before perfusion of 4% paraformaldehyde to increase the immunostaining of ghrelin neurons¹⁹.

In situ hybridization

An equimolar mixture of two 45-mer antisense oligonucleotide probes was used for *in situ* hybridization. The sequences of the antisense probes correspond to nucleotides, 90–134 and 421–465. Probes were 3'-end labelled with ³⁵S-dATP. *In situ* hybridization was carried out as described elsewhere²⁷.

Received 21 July; accepted 7 October 1999.

- Bowers, C. Y. Growth hormone-releasing peptide (GHRP). *Cell. Mol. Life Sci.* **54**, 1316–1329 (1998).
- Smith, R. G. *et al.* Peptidomimetic regulation of growth hormone secretion. *Endocrine Rev.* **18**, 621–645 (1997).
- Smith, R. G. *et al.* A nonpeptidyl growth hormone secretagogue. *Science* **260**, 1640–1643 (1993).
- Akman, M. S., Girard, M., O'Brien, L. F., Ho, A. K. & Chik, C. L. Mechanisms of action of a second generation growth hormone-releasing peptide (Ala-His-D-Nal-Ala-Trp-D-Phe-Lys-NH₂) in rat anterior pituitary cells. *Endocrinology* **132**, 1286–1291 (1993).
- Cheng, K., Chan, W. W. S., Barreto, A., Convey, E. M. & Smith, R. G. The synergistic effects of His-D-Trp-Ala-Trp-D-Phe-Lys-NH₂ on growth hormone (GH)-releasing factor-stimulated GH release and intracellular adenosine 3',5'-monophosphate accumulation in rat primary pituitary cell culture. *Endocrinology* **124**, 2791–2798 (1989).
- Howard, A. D. *et al.* A receptor in pituitary and hypothalamus that functions in growth hormone release. *Science* **273**, 974–977 (1996).
- McKee, K. K. *et al.* Molecular analysis of rat pituitary and hypothalamic growth hormone secretagogue receptors. *Mol. Endocrinol.* **11**, 415–423 (1997).
- Civelli, O. Functional genomics: the search for novel neurotransmitters and neuropeptides. *FEBS Lett.* **430**, 55–58 (1998).
- Meunier, J. C. *et al.* Isolation and structure of the endogenous agonist of opioid receptor-like ORL1 receptor. *Nature* **377**, 532–535 (1995).
- Reinscheid, R. K. *et al.* Orphanin FQ: a neuropeptide that activates an opioidlike G protein-coupled receptor. *Science* **270**, 792–794 (1995).
- Sakurai, T. *et al.* Orexins and orexin receptors: a family of hypothalamic neuropeptides and G protein-coupled receptors that regulate feeding behavior. *Cell* **92**, 573–585 (1998).
- Hinuma, S. *et al.* A prolactin-releasing peptide in the brain. *Nature* **393**, 272–276 (1998).
- Sudoh, T., Kangawa, K., Minamino, N. & Matsuo, H. A new natriuretic peptide in porcine brain. *Nature* **332**, 78–81 (1988).
- Kitamura, K. *et al.* Adrenomedullin: a novel hypotensive peptide isolated from human pheochromocytoma. *Biochem. Biophys. Res. Commun.* **192**, 553–560 (1993).
- Kozak, M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292 (1986).
- Heijne, V. G. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* **14**, 4683–4690 (1986).
- Nakayama, K. Furin: a mammalian subtilisin/Kex2p-like endoprotease involved in processing of a wide variety of precursor proteins. *Biochem. J.* **327**, 625–635 (1997).
- Kojima, M. *et al.* Cloning and characterization of a novel natriuretic peptide in frog (*Rana catesbeiana*). *J. Biol. Chem.* **269**, 13136–13140 (1994).
- Alonso, G. Effects of colchicine on the intraneuronal transport of secretory material prior to the axon: a morphofunctional study in hypothalamic neurosecretory neurons of the rat. *Brain Res.* **453**, 191–203 (1988).
- Guan, X. M. *et al.* Distribution of mRNA encoding the growth hormone secretagogue receptor in brain and peripheral tissues. *Mol. Brain Res.* **48**, 23–29 (1997).
- Bisi, G. *et al.* Acute cardiovascular and hormonal effects of GH and hexarelin, a synthetic GH-releasing peptide, in humans. *J. Endocrinol. Invest.* **22**, 266–272 (1999).

- Berghe, G. V. *et al.* Reactivation of pituitary hormone release and metabolic improvement by infusion of growth hormone-releasing peptide and thyrotropin-releasing hormone in patients with protracted critical illness. *J. Clin. Endocrinol. Metab.* **84**, 1311–1323 (1999).
- Schroeder, K. S. & Neagle, B. D. FLIPR: a new instrument for accurate, high throughput optical screening. *J. Biomol. Screening* **1**, 75–80 (1996).
- Coward, P. *et al.* Controlling signaling with a specifically designed Gi-coupled receptor. *Proc. Natl Acad. Sci. USA* **95**, 352–357 (1998).
- Jonathan, N. B., Peleg, E. & Hoefler, M. T. Optimization of culture conditions for short-term pituitary cell culture. *Methods Enzymol.* **103**, 249–257 (1983).
- Date, Y. *et al.* Orexin, orexigenic hypothalamic peptides, interact with autonomic, neuroendocrine and neuroregulatory systems. *Proc. Natl Acad. Sci. USA* **96**, 748–753 (1999).
- Date, Y. *et al.* Enterochromaffin-like cells, a cellular source of uroguanylin in rat stomach. *Endocrinology* **140**, 2398–2404 (1999).

Acknowledgements

We thank R. Haruno for technical assistance; H. Kaiya and T. Maki for technical advice; and R. J. Adams for comments on the manuscript. This work was supported in part by Special Coordination Funds for Promoting Science and Technology from the Science and Technology Agency (Encouragement System of COE) (to H.M.), and a grant-in-aid for the Promotion of Fundamental Studies in Health Science from the Organization for Pharmaceutical Safety and Research of Japan (to K.K.).

Correspondence and requests for materials should be addressed to K.K. (e-mail: kangawa@ri.ncvc.go.jp). The nucleotide sequence data reported here will appear in the DDBJ/EMBL/GenBank nucleotide sequence databases with the accession numbers AB029433 (rat) and AB029434 (human).

Central inputs mask multiple adult neural networks within a single embryonic network

Yves Le Feuvre, Valérie S. Fénelon & Pierre Meyrand

Laboratoire de Neurobiologie des Réseaux, CNRS et Université Bordeaux I, UMR 5816, Bâtiment B2, Avenue des Facultés, F-33405 Talence, France

It is usually assumed that, after construction of basic network architecture in embryos¹, immature networks undergo progressive maturation to acquire their adult properties^{2–4}. We examine this assumption in the context of the lobster stomatogastric nervous system. In the lobster, the neuronal population⁵ that will form this system is at first organized into a single embryonic network that generates a single rhythmic pattern⁶. The system then splits into different functional adult networks⁶ controlled by central descending systems^{7,8}; these adult networks produce multiple motor programmes, distinctively different from the single output of the embryonic network. We show here that the single embryonic network can produce multiple adult-like programmes. This occurs after the embryonic network is silenced by removal of central inputs, then pharmacologically stimulated to restore rhythmicity. Furthermore, restoration of the flow of descending information reversed the adult-like pattern to an embryonic pattern. This indicates that the embryonic network possesses the ability to express adult-like network characteristics, but descending information prevents it from doing so. Functional adult networks may therefore not necessarily be derived from progressive ontogenetic changes in networks themselves, but may result from maturation of descending systems that unmask pre-existing adult networks in an embryonic system.

The stomatogastric nervous system, STNS, of the embryonic (as well as the adult) lobster, *Homarus gammarus*, which controls the rhythmic motions of the stomodeum and foregut, respectively, consists of four interconnected ganglia, one of which is the stomatogastric ganglion (STG). By mid-embryonic development, the STG contains its adult complement of 30 identified neurons, most of them motor neurons⁵. In the adult, this population is organized into two different networks, the gastric and the pyloric, that are well

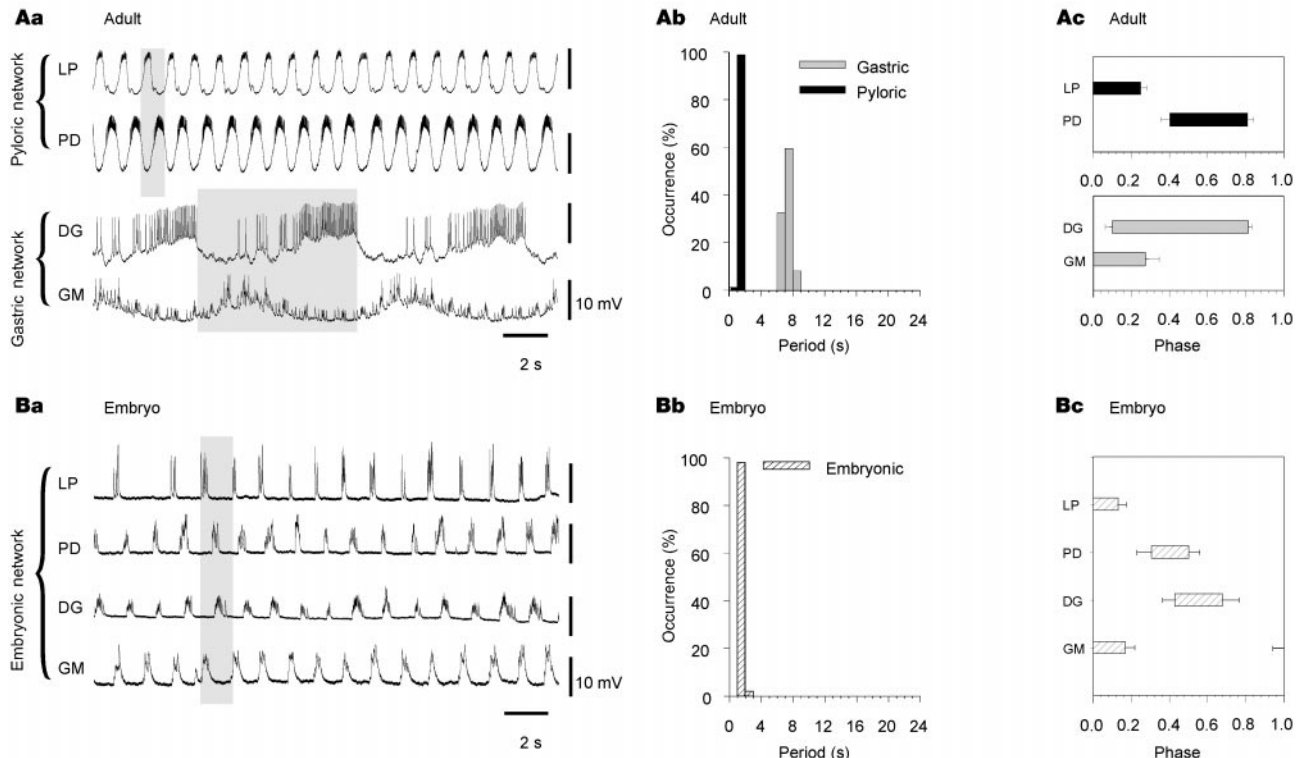


Figure 1 Spontaneous rhythmic activity generated by the same neurons in adult and embryonic stomatogastric nervous system of *H. gammarus*. **Aa**, Simultaneous recordings of the motor outputs generated by the adult pyloric (LP, PD) and gastric (DG, GM) STG neurons (shaded regions highlight the different periods of the two adult rhythms). **Ab**, Period repartition histogram of adult pyloric and gastric rhythms. **Ac**, Phase relationships of adult gastric (DG, GM) and pyloric (PD, LP) neurons. **Ba**, Simultaneous recordings of the motor output generated in the embryo by the same neurons as in **Aa**.

(Shaded region highlights the period of the single embryonic rhythm.) **Bb**, Period repartition histogram of embryonic rhythm. **Bc**, Phase relationships of embryonic neurons (DG, GM, PD and LP). Analysis of phase relationships for pyloric neurons (PD, LP) in the adult ($n = 3$) and the embryo ($n = 3$) shows significant statistical difference (Mann–Whitney rank sum test, $P < 0.001$) for the three parameters studied (end of LP discharge, beginning and end of PD discharge). DG, dorsal gastric; GM, gastric mill; LP, lateral pyloric; PD, pyloric dilator.

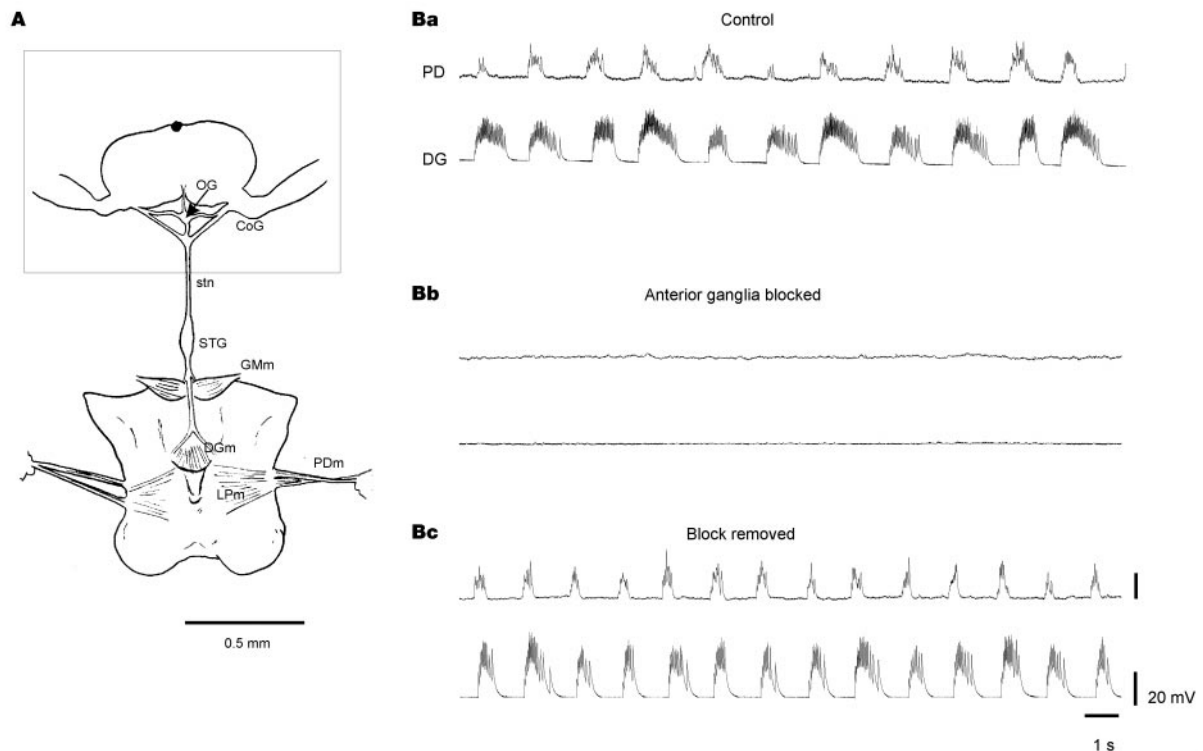


Figure 2 The spontaneous rhythmic activity of the embryonic network depends on the presence of descending inputs. **A**, Camera lucida drawing of the nerve–muscle embryonic preparation used for experiments. **B**, Gastric and pyloric network activities monitored in the presence (**a**, **c**) or in the absence (**b**) of central inputs. Suppression of

these inputs reversibly causes the embryonic network to fall silent. CoG, commissural ganglia; DGm, dorsal gastric muscle; GMm, gastric mill muscle; LPm, lateral pyloric muscle; OG, oesophageal ganglion; PDm, pyloric dilator muscle; STG, stomatogastric ganglion; stn, stomatogastric nerve.

characterized in terms of cellular and synaptic properties^{8,9}. *In vivo*, the two STG adult networks spontaneously and continuously generate two rhythmic motor patterns with different cycle periods (Fig. 1Aa and Ab)¹⁰. The pyloric period (two top traces, Fig. 1Aa) is stable for a given experiment, and ranges from 0.57 s to 1.38 s in

different preparations (mean, 0.95 ± 0.06 s; $n = 10$; all results are reported as mean $\pm$ s.e.m.). The distribution of pyloric period in one experiment is illustrated in Fig. 1Ab. Similarly, the gastric period (bottom two traces, Fig. 1Aa) ranges from 6.04 s to 17.86 s in different experiments (mean, 8.8 ± 1.09 s; $n = 10$). The distribution

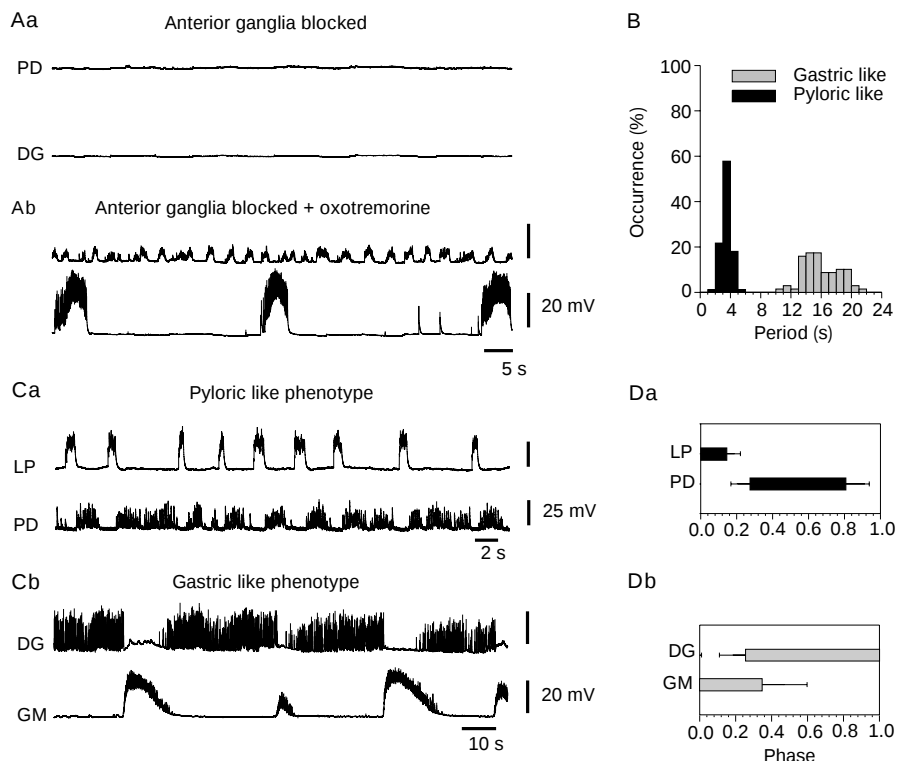


Figure 3 The embryonic stomatogastric network has the potential to express multiple adult-like rhythmic motor outputs. Removal of descending inputs results in complete silence of the embryonic network (Aa). Under these conditions, bath application of oxotremorine (10^{-5} M) elicits two different rhythmic activities (Ab), with periods similar to

those of adult pyloric (PD) and gastric (DG) rhythms (B). Furthermore, under these conditions, future pyloric and gastric neurons exhibit typical adult alternation (Ca, Cb), and express adult-like phase relationships (Da, Db).

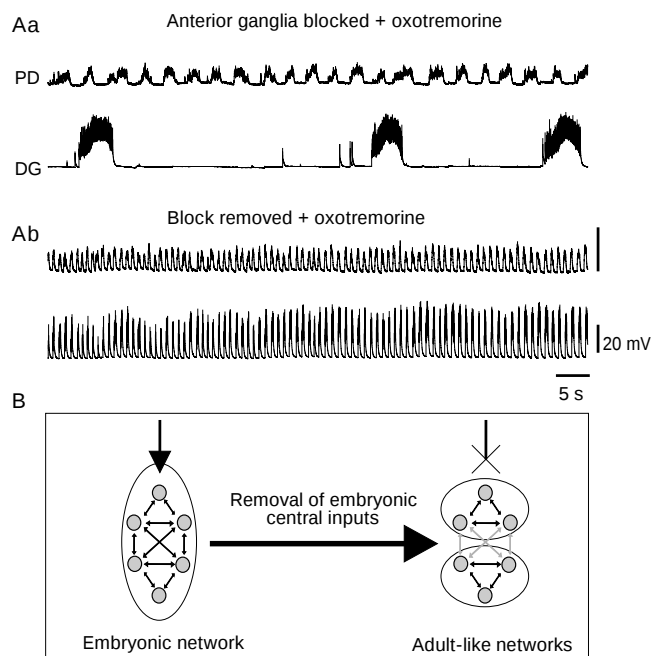


Figure 4 Descending inputs mask expression of multiple adult-like network activities in the embryonic STNS. Aa, After blockade of rostral STNS ganglia and pharmacological stimulation of embryonic STG neuronal population, embryonic STG neurons express two distinct adult-like outputs as illustrated in Fig. 3Ab. Ab, Restoration of the descending

inputs by removal of modified saline superfusion on anterior ganglia (see Fig. 2B) causes the network activity to switch from two adult-like rhythms to single embryonic rhythms. b, Adult-like network phenotypes are embedded in the embryonic network, but their expression is masked by central embryonic inputs.

of gastric periods in one experiment is illustrated in Fig. 1Ab. The STG neurons express stable phase relationships within their own network (see Fig. 1Ac and refs 11, 12). In contrast to the adult, the same STG neuronal population is organized in the embryo into a single network⁶ generating a single motor pattern with a period ranging from 0.92 s to 1.74 s (mean, 1.30 ± 0.07 s; $n = 10$; Fig. 1Ba and Bb). Moreover, the embryonic STG neurons all express stable phase relationships coordinated within the single embryonic rhythm (Fig. 1Bc). We have now investigated the mechanisms by which this ontogenetic switch—from a single functional network (Fig. 1Ba) to multiple adult networks (Fig. 1Aa)—may occur.

In the adult, the expression of STG networks is strictly dependent upon the descending instructions, including both long-lasting modulatory and transient synaptic effects, that arise from more rostral ganglia^{7,13} and travel through a single input nerve, the stomatogastric nerve (stn). As has been shown recently, in the embryo, the STG input nerve (Fig. 2A) contains a set of immature descending fibres^{5,14,15}. We tested the influence of these inputs on the expression of the embryonic STG network (Fig. 2Ba). When these inputs are reversibly suppressed by blocking the activity of anterior ganglia (see Methods) all embryonic STG neurons cease firing (Fig. 2Bb; $n = 8$), and therefore the network does not express any rhythmic motor output. After descending information is restored, the same neurons express rhythmic bursts of activity characteristic of the embryonic output (Fig. 2Bc). This result shows that, as in the adult^{7,13}, embryonic network activity is under the control of descending inputs. This indicates that at least some of the descending systems are functional early in development.

In the absence of descending inputs in the adult STNS, some modulatory substances are able to restore typical rhythmic motor patterns^{16–18}. Among the most potent substances to stimulate the adult quiescent STG are the muscarinic cholinergic agonists^{19,20}. To determine whether the muscarinic agonist, oxotremorine, is also able to restore the unique embryonic rhythmic motor output, we bath applied it to an embryonic quiescent preparation (anterior ganglia blocked, Fig. 3Aa). However, instead of restoring a typical embryonic network output, oxotremorine evoked two distinct rhythmic motor outputs (Fig. 3Ab; $n = 17$). The slower rhythm had a period ranging from 10.03 s to 52.14 s (mean, 21.99 ± 4.98 s; $n = 9$; see Fig. 3B), and the faster one had periods ranging from 2.04 s to 4.45 s (mean, 3.68 ± 0.23 s; $n = 9$; see Fig. 3B). The faster rhythm was expressed by a subset of neurons that will belong to the future pyloric network (Fig. 3Ca): these neurons expressed alternating activity similar to the adult (Fig. 3Da). The slower rhythm was found in future gastric motor neurons (Fig. 3Cb). They also expressed an alternating output pattern (Fig. 3Db). Although these two periods are different from those recorded in the adult *in vitro* STNS, they are similar to the adult pyloric and gastric periods recorded *in vivo* in unfed animals²¹.

Thus, in the embryo, the main features of adult networks—the ability to generate two separate patterns—are already found, but not expressed spontaneously. We then performed a series of experiments to find out whether central inputs are able to continuously and actively mask the expression of adult-like patterns in the embryo. In these experiments, all descending inputs were removed and pharmacological stimulation was provided to reveal adult-like phenotypes (Fig. 4Aa, as in Fig. 3Ab): we then restored the flow of endogenous descending information in the continued presence of pharmacological stimulation (Fig. 4Ab). In contrast to the adult system, where under similar conditions two different motor patterns would be maintained, the embryonic adult-like networks reverted immediately to their embryonic mode when the block was removed (Fig. 4Ab, $n = 10$).

These data show that as-yet unidentified descending inputs are important in maintaining the embryonic STG neuronal population as a unique functional network, while masking in a long-lasting fashion the adult-like potentialities of this embryonic population.

Our results indicate that adult networks can be present during embryonic life, but their expression is prevented by descending inputs (Fig. 4B). This developmental mechanism contrasts with already known ontogenetic mechanisms where there is a progressive maturation of an embryonic network involving a progressive acquisition of synaptic as well as intrinsic properties of their constituent elements^{2–4,22–29}. We conclude that the embryonic network must not necessarily be viewed as an immature structure, but can rather be seen as a hidden adult-like network. □

Methods

Experimental procedures

Dissection and electrophysiological recordings were performed as previously described in the adult⁸ and the embryo⁶ except that, for embryo, recording micropipettes were filled with 2 M potassium acetate (20–40 MΩ). Experiments on embryos were conducted with developmental stages ranging from E75 to E95 (for developmental staging see ref. 6). Spontaneous sequences of STG output were monitored intracellularly from motoneuron somata in the adult, and from their target stomodaeal muscles in the embryo. The activity of anterior ganglia containing descending neurons that project to the embryonic STG was reversibly blocked by superfusion of either modified saline (CholineCl 479.12 mM, KCl 12.74 mM, CaCl₂ 13.2 mM, MgSO₄ 3.9 mM, HEPES 5 mM) or isotonic sucrose solution (750 mM) in a small Vaseline well surrounding anterior ganglia (the rectangle in Fig. 2A indicates the area surrounded by the Vaseline well). Pharmacological stimulation was performed by bath application of oxotremorine (10^{-5} M, Sigma).

Analysis of data

Quantitative analysis was performed using spike2 software from Cambridge Electronic Design. Each cycle period histogram was computed from one experiment using 50–250 consecutive cycles. Phase relationships were computed from 20–180 consecutive cycles. The beginning and end of each box represent mean ($\pm$ s.e.) onset and offset phase of bursts in the indicated neurons. In the embryo, the onset of each burst was defined as an increase in instantaneous frequency of at least 5 Hz following a pause of at least 0.25 s. In both the embryo with the anterior ganglia blocked and the adult, the onset of each pyloric burst was defined as an increase in instantaneous frequency of at least 5 Hz following a silence of at least 0.5 and 1 s, respectively, while the onset of each gastric burst was defined as an increase in instantaneous frequency of at least 1 Hz following a pause of at least 2 s.

Received 23 July; accepted 28 October 1999.

- Fénelon, V. S., Casasnovas, B., Simmers, J. & Meyrand, P. Development of rhythmic pattern generators. *Curr. Opin. Neurobiol.* **8**, 705–709 (1998).
- Sillar, K. T. Synaptic specificity: development of locomotor rhythmicity. *Curr. Opin. Neurobiol.* **4**, 101–107 (1994).
- Spitzer, N. C. Development of voltage-dependent and ligand-gated channels in excitable membranes. *Prog. Brain Res.* **102**, 169–179 (1994).
- Levine, R. B., Morton, D. B. & Restifo, L. L. Remodeling of the insect nervous system. *Curr. Opin. Neurobiol.* **5**, 28–35 (1995).
- Fénelon, V. S., Casasnovas, B., Faumont, S. & Meyrand, P. Ontogenetic alteration in peptidergic expression within a stable neuronal population in lobster stomatogastric nervous system. *J. Comp. Neurol.* **399**, 289–305 (1998).
- Casasnovas, B. & Meyrand, P. Functional differentiation of adult neural circuits from a single embryonic network. *J. Neurosci.* **15**, 5703–5718 (1995).
- Moulins, M. & Courmil, I. All-or-none control of the bursting properties of the pacemaker neurons of the lobster pyloric pattern generator. *J. Neurobiol.* **5**, 447–458 (1982).
- Silverston, A. I. & Moulins, M. *The Crustacean Stomatogastric System* (Springer, Berlin, 1987).
- Harris-Warrick, R. M., Marder, E., Silverston, A. I. & Moulins, M. *Dynamic Biological Networks. The Stomatogastric Nervous System* (MIT Press, Cambridge, Massachusetts, 1992).
- Meyrand, P., Simmers, J. & Moulins, M. Construction of a pattern-generating circuit with neurons of different networks. *Nature* **351**, 60–63 (1991).
- Faumont, S., Simmers, J. & Meyrand, P. Activation of a lobster motor rhythm-generating network by disinhibition of permissive modulatory inputs. *J. Neurophysiol.* **80**, 2776–2780 (1998).
- Masabau, J.-C. & Meyrand, P. Modulation of a neural network by physiological levels of oxygen in lobster stomatogastric ganglion. *J. Neurosci.* **16**, 3950–3959 (1996).
- Robertson, R. M. & Moulins, M. Oscillatory command input to the motor pattern generators of the crustacean stomatogastric ganglion. II. The gastric rhythm. *J. Comp. Physiol. A* **154**, 473–491 (1984).
- Kilman, V. *et al.* Sequential developmental acquisition of cotransmitters in identified neurons of the stomatogastric nervous system of the lobsters, *Homarus americanus* and *Homarus gammarus*. *J. Comp. Neurol.* **408**, 318–334 (1999).
- Fénelon, V. S., Kilman, V., Myrand, P. & Marder, E. Sequential developmental acquisition of neuromodulatory inputs to a central pattern generating network. *J. Comp. Neurol.* **408**, 335–351 (1999).
- Hooper, S. L. & Marder, E. Modulation of a central pattern generator by two neuropeptides, proctolin and FMRFamide. *Brain Res.* **305**, 186–191 (1984).
- Flamm, R. E. & Harris-Warrick, R. M. Aminoergic modulation in lobster stomatogastric ganglion. I. Effects on motor pattern and activity of neurons within the pyloric circuit. *J. Neurophysiol.* **55**, 847–865 (1986).
- Turrigiano, G. G. & Silverston, A. I. Cholecystokinin-like peptide is a modulator of a crustacean central pattern generator. *J. Neurosci.* **9**, 2486–2501 (1989).
- Elson, R. C. & Silverston, A. I. Mechanisms of gastric rhythm generation in the isolated stomatogastric ganglion of spiny lobsters: bursting pacemaker potentials, synaptic interactions, and muscarinic modulation. *J. Neurophysiol.* **68**, 890–907 (1992).

20. Bal, T., Nagy, F. & Moulins, M. Muscarinic modulation of a pattern-generating network: control of neuronal properties. *J. Neurosci.* **14**, 3019–3035 (1994).
21. Clemens, S., Massabau, J.-C., Legeay, A., Meyrand, P. & Simmers, J. *In vivo* modulation of interacting central pattern generators in lobster stomatogastric ganglion: Influence of feeding and partial pressure of oxygen. *J. Neurosci.* **18**, 2788–2799 (1998).
22. Bolshakov, V. Y. & Siegelbaum, S. A. Regulation of hippocampal release during development and long-term potentiation. *Science* **269**, 1730–1734 (1995).
23. McCormick, D. A., Trent, F. & Ramoa, A. S. Postnatal development of synchronized network oscillations in the ferret dorsal lateral geniculate and perigeniculate nuclei. *J. Neurosci.* **15**, 5739–5752 (1995).
24. Durand, G. M., Kovalchuk, Y. & Konnerth, A. Long-term potentiation and functional synapse induction in developing hippocampus. *Nature* **381**, 71–75 (1996).
25. Katz, L. C. & Shatz, C. J. Synaptic activity and the construction of cortical circuits. *Science* **274**, 1133–1138 (1996).
26. Nick, T. A., Kaczmarek, L. K. & Carew, T. J. Ionic currents underlying developmental regulation of repetitive firing in *Aplysia* bag cell neurons. *J. Neurosci.* **16**, 7583–7598 (1996).
27. Warren, R. A. & Jones, E. G. Maturation of neuronal form and function in a mouse thalamo-cortical circuit. *J. Neurosci.* **17**, 277–295 (1997).
28. Sun, Q.-Q. & Dale, N. Developmental changes in expression of ion currents accompany maturation of locomotor pattern in frog tadpoles. *J. Physiol. (Lond.)* **507**, 257–264 (1998).
29. Angulo, M. C., Staiger, J. F., Rossier, J. & Audinat, E. Developmental synaptic changes increase the range of integrative capabilities of an identified excitatory neocortical connection. *J. Neurosci.* **19**, 1566–1576 (1999).

Acknowledgements

We thank T. Bem, E. Marder, R. Miles, J. Simmers and S. Faumont for comments on an earlier version of the manuscript, and S. Faumont for providing the adult intracellular recordings shown in Fig. 1A.

Correspondence and requests for materials should be addressed to P.M.
(e-mail: p.meyrand@lnr.u-bordeaux.fr).

Short-term memory in olfactory network dynamics

Mark Stopfer & Gilles Laurent

Division of Biology, 139-74, California Institute of Technology, Pasadena, California 91125, USA

Neural assemblies in a number of animal species display self-organized, synchronized oscillations in response to sensory stimuli in a variety of brain areas.^{1–5} In the olfactory system of insects, odour-evoked oscillatory synchronization of antennal lobe projection neurons (PNs) is superimposed on slower and stimulus-specific temporal activity patterns. Hence, each odour activates a specific and dynamic projection neuron assembly whose evolution during a stimulus is locked to the oscillation clock^{6,7}. Here we examine, using locusts, the changes in population dynamics of projection-neuron assemblies over repeated odour stimulations, as would occur when an animal first encounters and then repeatedly samples an odour for identification or localization. We find that the responses of these assemblies rapidly decrease in intensity, while they show a marked increase in spike time precision and inter-neuronal oscillatory coherence. Once established, this enhanced precision in the representation endures for several minutes. This change is stimulus-specific, and depends on events within the antennal lobe circuits, independent of olfactory receptor adaptation: it may thus constitute a form of sensory memory. Our results suggest that this progressive change in olfactory network dynamics serves to converge, over repeated odour samplings, on a more precise and readily classifiable odour representation, using relational information contained across neural assemblies.

Natural olfactory behaviour is generally characterized by repeated odour samplings. This feature is imposed by olfactory physiology—breathing-coupled sniffing in vertebrates^{8,9} or olfactory appendage flicking in arthropods¹⁰—as well as by the complex physical nature

of odour plumes, which contain intermixed odorized and non-odorized filaments¹¹. We thus considered the possibility that successive odour samples might be accompanied by enduring changes in their representation by the central nervous system, other than ones simply explained by receptor adaptation. In particular, we examined the possibility that the dynamical or temporal features of odour representation which we characterized previously^{6,7} might themselves evolve as the animal becomes more familiar with an

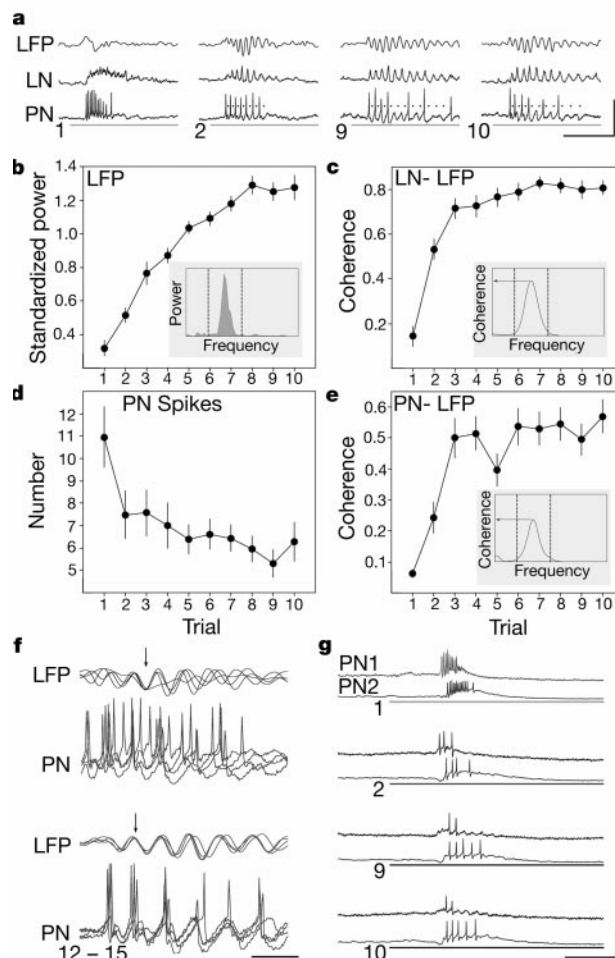


Figure 1 Response intensity decreases, while coherence and spike time precision increase, over repeated odour presentations. **a**, Simultaneous local field potential (LFP) and intracellular recordings from a local neuron (LN) and projection neuron (PN). Early (1–2) and late (9–10) trials are shown (dots: LN-caused IPSPs). LFPs were bandpass-filtered (5–55 Hz) to emphasize 20–30 Hz odour-elicited oscillations. Horizontal bar indicates odour delivery. Calibration: horizontal, 300 ms; vertical, 0.8 (LFP), 10 (LN), 40 (PN) mV. **b**, LFP power spectrum (see Methods) increased during the first 6 or 7 trials before reaching asymptote. Power measures for each experiment (mean $\pm$ s.e.m., $n = 40$) were standardized by each experiment's mean response. 2-way ANOVA: $f_{\text{trials}}(9) = 42.9$, $P < 0.0001$. Inset: example power spectrum, 5–55 Hz; dotted lines enclose the integrated band. **c**, Coherence between LNs and LFP (see Methods and inset) increased rapidly over stimulus trials. $n = 27$; 2-way ANOVA: $f_{\text{trials}}(9) = 50.82$, $P < 0.0001$. **d**, The number of odour-elicited action potentials in PNs decreased markedly by the second trial. All spikes within the ~ 1 -s odour response period were counted. $n = 23$; 2-way ANOVA: $f_{\text{trials}}(9) = 8.62$, $P < 0.0001$. **e**, Coherence between PN spike time and LFP increased over trials; $n = 22$; 2-way ANOVA: $f_{\text{trials}}(9) = 17.18$, $P < 0.0001$. **f**, Odour-elicited PN action potentials became increasingly locked to the LFP. Superimposed traces (all from the same experiment, different from that in **a**) aligned on the first LFP cycle (arrow). Horizontal calibration: 50 ms. Vertical scale is standardized to emphasize timing. **g**, Odour-elicited responses in two simultaneously recorded PNs. Calibration: horizontal, 200 ms; vertical, 70/40 mV.

odour. Such an evolution would constitute a form of unsupervised or non-associative plasticity.

We delivered a series of identical 1-second-long odour puffs at 0.1 Hz to one antenna of a locust, while recording simultaneously intracellular potentials from local and projection neurons (LNs and PNs) in the ipsilateral antennal lobe, and extracellular local field potentials (LFPs) from the ipsilateral mushroom body (Fig. 1a). The stimulus sequence was delivered to an initially naive animal, that is, one that had no prior experience with this odour. The first of the stimuli typically elicited very strong LN and PN responses (from only those LNs and PNs tuned to the odour presented), with peak PN instantaneous firing rates greater than 20–30 Hz but with no detectable periodic subthreshold or spiking activity at this frequency in either LNs or PNs (Fig. 1a, f, g). Similarly, the LFP waveform contained very little power at ~20 Hz (Fig. 1a, b), a feature normally characteristic of odour-evoked LFPs in animals already familiar with the odour tested^{12,13}.

Over the course of the succeeding stimuli, however, two concurrent phenomena developed. First, the response intensity of PNs

(total number of spikes) decreased markedly between trials 1 and 2 and a little more over the following trials (Fig. 1a, d). Second, 20 ± 5 Hz periodic activity appeared at trials 2 or 3 and progressively developed, as seen clearly in LNs and LFPs (Fig. 1a), but also in PNs (Fig. 1a, f). The power of the extracellular LFP oscillatory activity around 20 Hz increased greatly between trials 1 and 8, reaching a maximum that was unaffected by further odour stimulation (Fig. 1b). The coherence (real value at the peak oscillation frequency) calculated both between LN membrane potential and LFP waveform and between PN spike time and LFP waveform increased 4–5 times over these trials (Fig. 1c, e). Because PN spiking activity—which causes the LFP recorded in their target area¹³—decreased while oscillatory power increased, we conclude that PN spikes became more precisely timed with respect to one another as the odour became more familiar. This can be seen directly by inspection of superimposed PN/LFP traces taken from sets of early and late trials (Fig. 1f), and of paired PN recordings over the first 10 odour trials (Fig. 1g). Note the precise relative patterning of the spikes evoked in these PNs in trials 9 and 10 (Fig. 1g), as previously shown to contain odour-specific information⁷. Such increased precision in patterning was always accompanied by an enhancement of the LN-evoked inhibitory postsynaptic potentials¹⁴ (IPSPs) in PNs (Fig. 1a, g).

This response evolution was independent of the interstimulus interval (range: 2.5–20 s, Fig. 2a) and of the duration of odour pulses (range: 0.25–2 s, Fig. 2b). Series of odour pulses of random durations and interstimulus intervals, mimicking natural odour stimulation by airborne plumes¹¹, also caused the same evolution (Fig. 2c). The development of synchronization within these circuits could occur during the first trial also if odour delivery was maintained for several seconds during that trial (>2 s, *n* = 6). However, intermittent stimuli were generally more effective (Fig. 2d; a statistical analysis is available; see Supplementary Information), possibly because prolonged stimuli induced receptor adaptation. Repeated or prolonged experience with an odour thus progressively led to a less intense (that is, containing fewer spikes) but more precise central representation in the antennal lobe, caused in part at least by the growing strength or efficacy of periodic inhibition by LNs.

Once established by repeated stimulation with one odour, this state change persisted for several minutes in the absence of intervening stimulation with that same odour. The influence of the duration of a pause in odour stimulation after trial 10 on oscillatory synchronization at later trials is plotted in Fig. 3a. An interval of 12 min, for example, appeared sufficient to reset the system to a naive state. This memory is therefore a short-term form. If, following a long pause (>15 min), stimulation with the same odour was resumed, the response pattern expressed by an individual PN or LN during the second series of trials followed an evolution similar to that seen during the first series.

This experience-related state change was odour-specific, showing no carry-over across responses caused by chemically distinct odours (for example, aliphatic alcohols to terpenes ('diff') or apple to cherry blends ('new'), Fig. 3b). Interpolating a novel odour within a sequence of stimulus with a familiar odour did not affect the response to the familiar odour (not shown). Exposure to an odour, however, increased the probability of oscillatory synchronization upon the first presentation of a chemically related odour. Such carry-over occurred best across related aliphatic alcohols (for example, hexanol to octanol, or vice versa; labelled 'sim') and to a lesser extent, from odours to binary blends containing them (odour A to blend AB; labelled A → AB) (Fig. 3b).

These results suggest that the observed state changes upon repeated presentation of an odour are not global, but restricted to those neurons and/or synapses activated by that odour; they do not, however, indicate whether these changes are peripheral, central or both. A simple explanation could be that receptor adaptation over

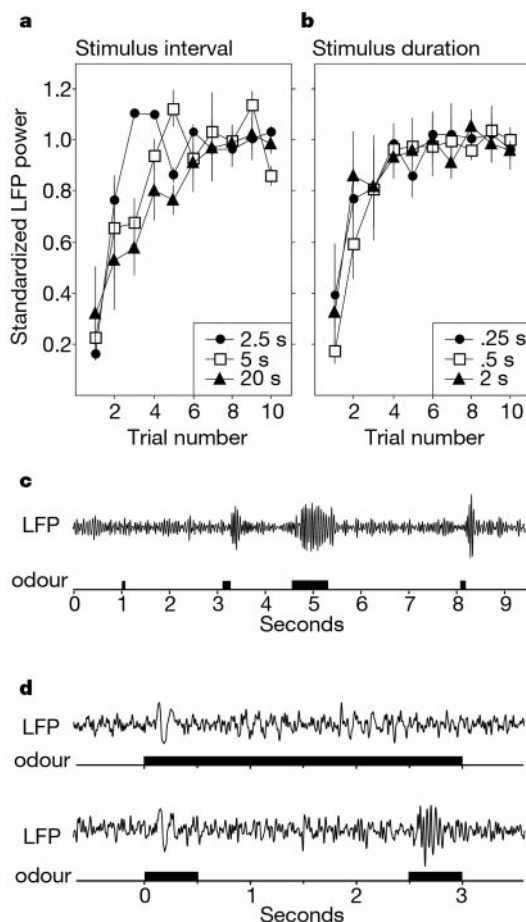


Figure 2 Response evolution is caused by a wide variety of stimulation regimes. **a**, Odour puffs delivered at 3 intervals (2.5 s, *n* = 2; 5 s, *n* = 4; 20 s, *n* = 4), and **b**, for 3 puff durations (0.25 s, *n* = 2; 0.5 s, *n* = 4; 2 s, *n* = 3) all evoked increases in LFP power similar to those observed for 0.1 Hz, 1 s stimulation (Fig. 1). Results from each experiment were standardized by the mean of trials 8–10. 1-way ANOVA: for each condition there was a significant effect of trials (*P* < 0.05). **c**, Irregular patterns of odour pulse duration and interstimulus interval also caused the same evolution (*n* = 6, typical example shown, LFP bandpass-filtered (15–35 Hz); the nature of the experiment—confounded duration and interval variables—precluded quantifying these results as in **a** and **b**). **d**, In this example, a single 3-s odour puff did not elicit strong oscillations (top LFP). Two shorter pulses, spaced so as to occur within the same 3-s period, did evoke strong oscillations upon the second stimulation (bottom LFP). Top and bottom recordings are 15 min apart. LFPs were bandpass-filtered (5–55 Hz).

successive trials causes the intensity of the peripheral (antennal) output to decrease, and progressively enter a range of intensities appropriate for evoking oscillatory dynamics in the antennal lobe¹⁵. Because olfactory receptor neurons are distributed over the length of the antenna^{16,17}, we could test this hypothesis directly. We 'trained' animals with an odour repeatedly presented to the receptors on one part of the antenna, and subsequently tested whether the same odour, this time presented to the non-adapted receptors on the other part of the antenna, would immediately cause oscillatory synchronization of antennal lobe neurons. Stimulation of non-

adapted receptors always gave rise to a strongly synchronized response in a 'trained' antennal lobe (Fig. 4). Thus, the state changes caused by experience with an odour depend, at least in part, on central and stimulus-specific changes within the antennal lobe. In addition, these results indicate that activation of non-adapted receptors does not preclude the activation of oscillatory dynamics within antennal lobe networks, provided these networks are not naive.

How might evolving network dynamics contribute to odour identification? Odour-specific relational information, previously shown to be contained in spike sequences read within and between PN⁷, cannot be obtained from the first trial of a series if the trial is too short, and becomes increasingly reliable as the number of successive experiences with that odour increases. This suggests that responses to later trials should permit more reliable odour identification. However, because the odour-specific firing patterns in PN responses seem to be most distinct when the response intensity is greatest (Fig. 5a), the responses of a PN to two odours would appear to be most reliably discriminated from one another at trial 1. If each successive sample evokes a different representation, which one of them—the first, most salient one, or a later, stable and more precise one—should the animal rely on for storage, or for comparison to a stored template, or to another suddenly occurring odour?

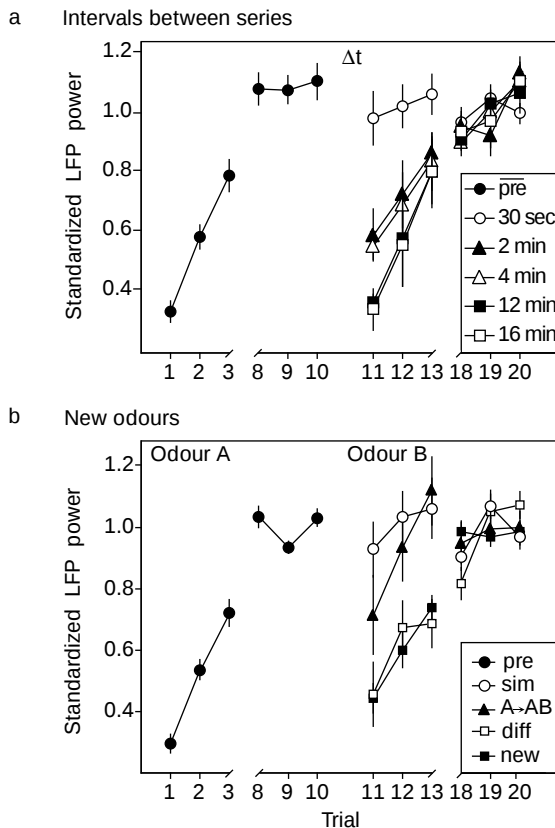


Figure 3 The tendency to oscillate in response to an odour endures for several minutes and is odour-specific. **a**, As a pre-test, an odorant was delivered (1-s puff) 10 times (0.1 Hz, pre = mean of all series), resulting in a significant increase in LFP power (2-way ANOVA: $f_{\text{trials}}(5) = 33.53$, $P < 0.0001$). Then, odour delivery was interrupted for an interval Δt , after which, as the post-test, stimuli resumed (trials 11–20). Results from each experiment were standardized by the mean of trials 18–20. Interval effects were similar for all odorants tested. Numbers of experiments: 30 s, 32; 2 min, 12; 4 min, 7; 12 min, 4; 16 min, 5. A 2-way ANOVA revealed a significant effect of interval over trials: $f_{\text{interval} \times \text{trials}}(20) = 2.16$, $P < 0.004$; at the first post-test trial 11 interval groups responded differently from each other (1-way ANOVA: $f(4, 59) = 5.42$, $P < 0.0009$). The 30-s interval group at trial 11 was not different from trial 10 (1-way ANOVA: $f(1, 63) = 1.17$ ns; the 12-min and 16-min interval groups at trial 11 were not different from baseline, trial 1 (2-way ANOVA: $f_{\text{interval} \times \text{trial}}(1) = 0.01$, n.s.). **b**, Pre-test trials were made as in **a**, and caused a significant increase in LFP power (2-way ANOVA: $f_{\text{trials}}(5) = 78.16$, $P < 0.0001$). Odorants were changed (following a 30-s pause to discharge the odour delivery system, right side of graph) to 'sim', ($n = 9$), a molecularly similar odorant; 'A $\rightarrow$ AB', ($n = 7$), a blend (AB) of which the initial odorant (A) was a component; 'diff', ($n = 4$), a molecularly different odorant; 'new', ($n = 5$), a new compound odorant. These second odour groups responded differently from each other over trials (2-way ANOVA: $f_{\text{odour} \times \text{trial}}(15) = 3.78$, $P < 0.0001$; at the first post-test trial, the group responses were significantly different from each other (1-way ANOVA: $f(3, 73) = 6.32$, $P < 0.0007$). For each experiment, results with odours A and B were standardized by the mean of responses at trials 8–10 and 18–20, respectively.

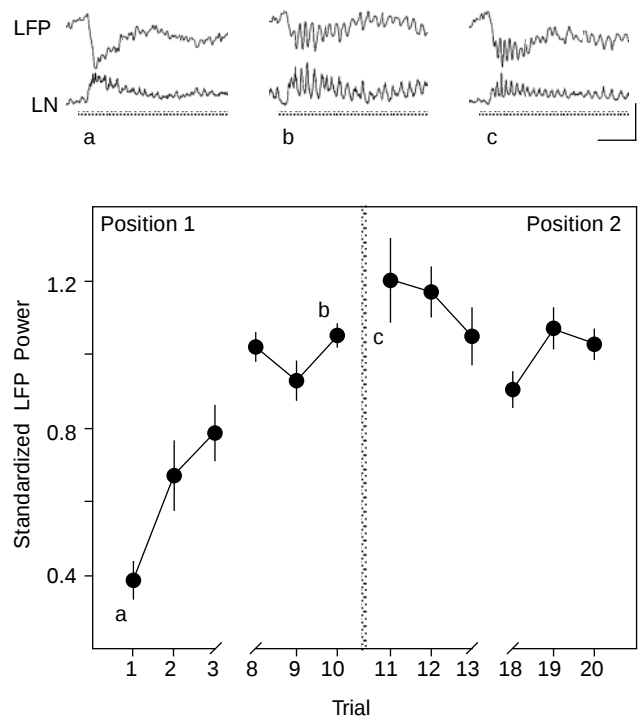


Figure 4 The evolution in antennal lobe dynamics depends on changes within the antennal lobe. A barrier separated the olfactory receptor neurons of the antenna into two sets (positions 1 and 2). Ten stimuli were delivered with the odour pipette in position 1 (1 s, 0.1 Hz), eliciting LFP oscillations that grew from weak (trial 1, **a**) to strong (trial 10, **b**); 2-way ANOVA: $f_{\text{trials}}(5) = 20.05$, $P < 0.0001$. The odour pipette was then moved to position 2; after a 30-s period to discharge the delivery apparatus, stimuli resumed. The stimuli immediately elicited strong oscillations (**c**); a 2-way ANOVA revealed a significant effect of stimulus position over trials: $f_{\text{position} \times \text{trials}}(5, 5) = 13.89$, $P < 0.0001$, and responses at trial 11 were not significantly different from those at trial 10 (t -test: $t(20) = 1.29$, n.s.). Results for each experiment ($n = 11$) were standardized by the mean of responses at the last 3 trials or each position. Odorants did not 'leak' around the barrier: delivering stimuli to a Vaseline-coated portion of the antenna on one side of the barrier never evoked central responses. Calibration: horizontal, 250 ms; vertical, 0.6/6 mV.

We addressed this question by measuring an observer's chances of correct odour identification given the spike trains evoked in several PNs by two odours (A and B) over 10 successive trials with each (Fig. 5a). We ignored here all relational information, and relied exclusively on odour-specific information contained in individual PN spike trains, assessed by an intensity- and pattern-sensitive multidimensional clustering algorithm¹⁸. This algorithm maps each spike train onto a point in a multidimensional space (see Methods). The relatedness of spike trains can thus be quantified by the distance separating the points which represent them. To determine how the evolution in PN response intensity and patterning might contribute to improvements or decrements in odour categorization, we used two strategies. In the first, we used trial 1 responses (a_1 and b_1) to odours A and B as the templates to which all successive responses ($a_{2...10}$, $b_{2...10}$) were compared (distance measures α and Δ respectively, Fig. 5b). Response a_4 of one PN to odour A at trial 4, for example, was then compared to a_1 (distance α) and b_1 (distance Δ), and classified as A if $\Delta > \alpha$ (Fig. 5b, c). Over all PNs and trials, 22% of PN responses misclassified the odours (Fig. 5d), and error rates, by definition nil at trial 1, increased with time. Hence, trial 1 templates are only reliable if sampling is not repeated. In the second strategy, we used the stabilized response patterns (centroid of a_{8-10} or b_{8-10}) as templates (Fig. 5e). In this case, only 4% of PN responses misclassified the odours over all trials (Fig. 5f, g). In

addition, the error rate decreased from a maximum of 18% at trial 1 to 0% by trial 5. Thus, although instantaneous firing rates are highest at trial 1 (naive state), odour categorization improves as response intensity decreases. Moreover, the repeated-sampling strategy ensures convergence to the correct classification. So when multiple odour samples occur, the use of less intense but stationary PN response patterns enables better odour classification.

In summary, we found that sequential exposure to an odour in the absence of reinforcer leads to rapid and significant short-term changes in the central representation of that odour by antennal lobe neurons. These changes include a decrease in PN activity, a development of oscillatory synchronization of PNs and LNs, and an increase in the inter-PN oscillatory coherence and spike time precision. These changes are odour-specific, with some transfer across the responses caused by chemically related stimuli, implying that similar odours activate separate, but overlapping, LN/PN assemblies and that this form of plasticity involves central, circuit-specific modifications. The effects last for several minutes and involve an apparent enhancement of periodic inhibition by LNs, leading to a specific temporal refinement of the population representation.

Our results have two main functional and practical implications. First, they suggest that accurate encoding of odour identity in the first olfactory relay should rely increasingly on distributed and

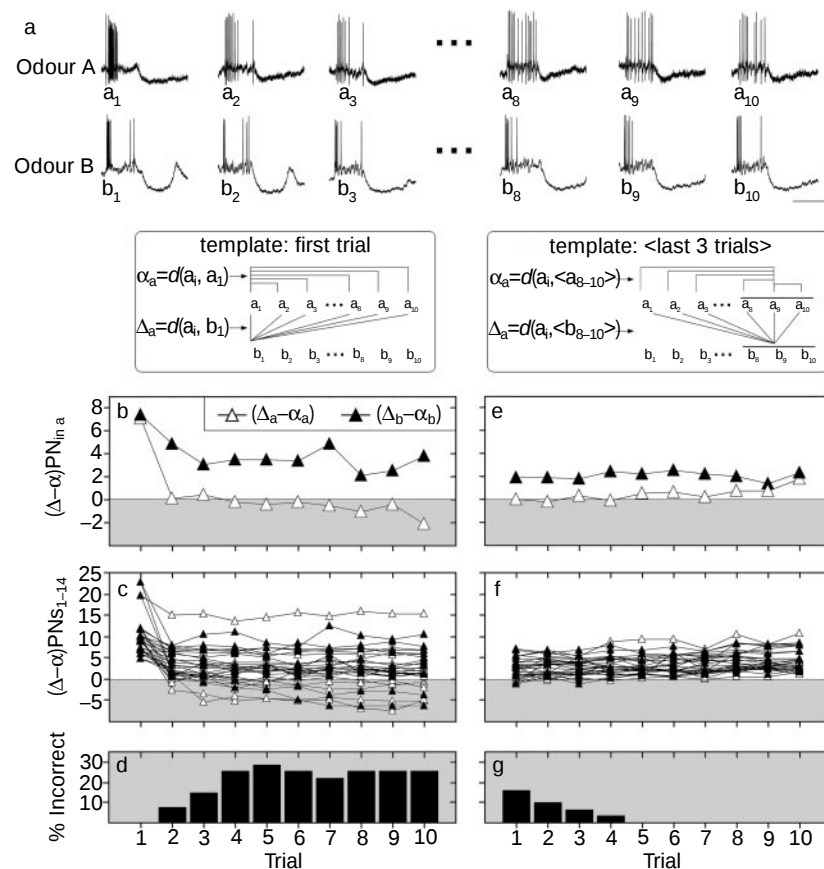


Figure 5 Individual PN responses to odours evolve over repeated trials, leading to more accurate odour classification. **a**, Intracellular recordings from a single PN, showing early-trial (1–3) and late-trial (8–10) responses to two odours, A and B. Odour-specific slow temporal patterns appear to be in place immediately, but other aspects of the response develop over trials. Calibration: horizontal, 1 s; vertical, 43 mV. **b–d**, Odours classified using the first response as a template. Inset: each PN response was compared to the first response elicited by that odour (distance α), and also to the first response elicited by the other odour (distance Δ). (Comparisons for odour A (α_a and Δ_a) are shown, but comparisons for odour B were made as well.) **b**, Plot of $(\Delta - \alpha)$ as a function of trial

number for PN in **a**. ($\Delta - \alpha$) is calculated separately for odours A (open triangles) and B (filled triangles). A negative value of $(\Delta - \alpha)$ indicates that the distance (a_i, a_1) is greater than the distance (a_i, b_1) and thus determines that a_i should be classified as caused by odour B at that trial. It is thus counted as 'incorrect', and used for the plot in **d**. **c**, As **b** but for all 14 PNs. **d**, Plot of percentage incorrect classification as a function of trial number, taken from data in **c**. **e–g**, Odours classified using the late, stabilized response (mean of the last 3 responses) as a template; using individual stabilized responses (trials 8, 9 or 10) as templates yielded the same results. Conventions as in **b–d**. **g**, Percentage of incorrect PN classifications.

relational features of population activity as an odour becomes familiar. Previous work showed that information about odour identity lies in the precise relationships of spike sequences in and across co-activated PNs^{6,7,18,19}. We show here that such relational information can be obtained only when PN activity is lowest, periodic and precise, that is, after repeated or prolonged exposures to the odour. Further, ignoring such relational information, we also showed that PN responses to two odours are most reliably classified for responses obtained after repeated samples of each odour, that is, when PN activity and instantaneous firing rates are lowest—and when stable responses are used as templates. Both lines of evidence indicate that repeated odour sampling should improve performance in categorization. While human psychophysical studies show that the first experience with an odour is reliably perceived as the most intense²⁰, we are unaware of evidence that accuracy decreases or increases with trial number. Behavioural data from both vertebrates and invertebrates show that stimulation with an odour usually triggers rapid and repeated sampling of that odour. The initial, intense response to odours may underlie detection and rapid but coarse identification; later responses may provide a more detailed characterization. Thus, we propose that repeated sampling serves to converge on a precise stimulus identification by exploiting the short-term plasticity of circuit dynamical behaviour. Such plasticity may provide the dynamical substrate for a sensory memory, permitting the early and rapid self-organization of neural responses to sensory input.

Second, our results illustrate how the dynamics of a sensory circuit can depend critically on its recent history. This is important for the interpretation of electrophysiological experiments. Sensory physiologists, in the course of an experiment, routinely look for one or several neurons and then search for an appropriate stimulus, or repeatedly provide a stimulus while searching for appropriate neurons. These seemingly innocuous searches can, as shown here, alter the responses not only of the neuron(s) tested, but also of others whose activity is not monitored, and can thereby alter population dynamics. Methods that require signal averaging over successive trials for enhancement of the signal-to-noise ratio, or that use large sets of shuffled and non-repeated stimuli, would respectively either fail to detect, or fail to cause, such an evolution. From the experimenter's perspective, repeated sampling is a tool that assumes stationarity. From the brain's perspective, repeated sampling may rather, under natural conditions, exploit the advantages of non-stationarity for improved performance, using increased response precision rather than increased response strength. □

Methods

Specimens

Results were obtained from 101 locusts (*Schistocerca americana*) from a crowded colony. Young adults of either sex were immobilized with one antenna intact; the brain was exposed, desheathed, and continually superfused with oxygenated locust saline, as previously described^{12,13}. In one experiment ($n = 7$ locusts), a barrier formed of a thin plastic card with a central hole was threaded part way along the antenna and sealed there with Vaseline, allowing separate stimulation of proximal and distal olfactory receptor arrays.

Odour stimulation

Odorant puffs (0.31 min^{-1}), carried by desiccated and filtered air, were delivered by individual pipettes (1-cm inner diameter) placed 2–3 cm in front of the antenna. A 10-cm diameter vacuum funnel placed 5 cm behind the antenna constantly drew background air over the antenna and rapidly vented odours. Each odour (2–3 μl apple, strawberry (Gilberties), cherry (Bell Flavors and Fragrances), spearmint (Flavco), eugenol, geraniol, 1-pentanol, 1-hexanol, 1-octanol (Sigma), cineole, isoamyl acetate, citral (Aldrich)), applied to a small strip of filter paper, was carried by a separate pipette. Before delivering the first in a series of odour puffs to the antenna, the pipette's standing content was repeatedly discharged into a separate vacuum funnel to ensure consistent odour concentrations at all trials.

Electrophysiology

LFPs were recorded using saline-filled blunt glass micropipettes (tip, $\sim 10 \mu\text{m}$, 3–7 M Ω), and amplified with a d.c. amplifier (NPI, Adams-List). Intracellular recordings from

antennal lobe neurons (including 200 paired intracellular recordings in 51 animals) were made using sharp glass micropipettes (150–250 M Ω , Sutter P87 horizontal puller) filled with 0.5 M potassium acetate and amplified with a separate d.c. amplifier (Axon Instruments). Data were stored on digital audio tape (DAT 8 channel, 5-kHz sampling/channel, Micro Data) and analysed off-line using NBM116L hardware, LabVIEW (National Instruments) and MatLab (The MathWorks) software. Non-phase shifting, band-pass filtering (1–55 Hz or 5–55 Hz, 5-pole, Butterworth) of LFPs was accomplished by a software algorithm.

Data analysis

Statistical comparisons were made by 1- and 2-way analyses of variance and *t*-tests, with significance level set at $P < 0.05$. LFP power spectra were measured from unfiltered LFPs. Average power for odour responses (Figs 1b; 2a, b; 3a, b; 4) was obtained by integrating a 15-Hz band centred on each preparation's peak odour response frequency (see inset, Fig. 1b). Coherence measures (amplitude only) were made using a multitaper method²¹ implemented by MatLab. LN waveforms were directly compared to the corresponding 5–55-Hz bandpass-filtered LFPs. PN rasters were convolved with a 15-ms gaussian; the resulting continuous waveforms were compared to the corresponding 5–55-Hz bandpass-filtered LFPs. Peak coherence within a 15-Hz band centred on each preparation's peak odour response frequency was measured (Fig. 1c, d). Similarities between odour-evoked PN spike trains were quantified by calculating the euclidean distance between spike trains analysed as previously described¹⁸. Briefly, each spike train was divided into n non-overlapping bins, and spikes in each bin were counted. The vector constructed from this list of numbers defined a point in n -space. The euclidean distances α and Δ between any two such points were then calculated (Fig. 5b, e). Bin sizes between 10 and 200 ms were used; all gave similar results. Results in Fig. 5 are with 100-ms bins over 3 s (30 bins, hence $n = 30$).

Received 23 August; accepted 22 October 1999.

- Adrian, E. D. Olfactory reactions in the brain of the hedgehog. *J. Physiol. (Lond.)* **100**, 459–473 (1942).
- Eckhorn, R. et al. Coherent oscillations: A mechanism of feature linking in the visual system? *Biol. Cybern.* **60**, 121–130 (1988).
- Gray, C. M., Koenig, P., Engel, A. K. & Singer, W. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* **338**, 334–337 (1989).
- Gelperin, A. & Tank, D. W. Odour-modulated collective network oscillations of olfactory interneurons in a terrestrial mollusc. *Nature* **345**, 437–440 (1990).
- Prechtl, J. C. Visual motion induces synchronous oscillations in turtle visual cortex. *Proc. Natl Acad. Sci. USA* **91**, 12467–12471 (1994).
- Laurent, G., Wehr, M. & Davidowitz, H. Temporal representations of odors in an olfactory network. *J. Neurosci.* **16**, 3837–3847 (1996).
- Wehr, M. & Laurent, G. Odour encoding by temporal sequences of firing in oscillating neural assemblies. *Nature* **384**, 162–166 (1996).
- Freeman, W. J. Spatial properties of an EEG event in the olfactory bulb and cortex. *Electroencephalogr. Clin. Neurophysiol.* **44**, 586–605 (1978).
- Gray, C. M. & Skinner, J. E. Centrifugal regulation of neuronal activity in the olfactory bulb of the waking rabbit as revealed by reversible cryogenic blockade. *Exp. Brain Res.* **69**, 378–386 (1988).
- Mellon, D. Physiological characterization of antennular flicking reflexes in the crayfish. *J. Comp. Physiol. A* **180**, 553–565 (1997).
- Murlis, J., Elkington, J. S. & Cardé, R. T. Odor plumes and how insects use them. *Annu. Rev. Entomol.* **37**, 505–532 (1992).
- Laurent, G. & Davidowitz, H. Encoding of olfactory information with oscillating neural assemblies. *Science* **265**, 1872–1875 (1994).
- Laurent, G. & Naraghi, M. Odorant-induced oscillations in the mushroom bodies of the locust. *J. Neurosci.* **14**, 2993–3004 (1994).
- MacLeod, K. & Laurent, G. Distinct mechanisms for synchronization and temporal patterning of odor-encoding neural assemblies. *Science* **274**, 976–979 (1996).
- Wehr, M. & Laurent, G. Relationship between afferent and central temporal patterns in the locust olfactory system. *J. Neurosci.* **19**, 381–390 (1999).
- Masson, C. & Mustaparta, H. Chemical information processing in the olfactory system of insects. *Physiol. Rev.* **70**, 199–245 (1990).
- Slifer, E. H. & Sekhon, S. S. The dendrites of the thin-walled olfactory pegs of the grasshopper (Orthoptera, Acrididae). *J. Morphol.* **114**, 393–410 (1965).
- MacLeod, K., Bäcker, A. & Laurent, G. Who reads temporal information contained across synchronized and oscillatory spike trains? *Nature* **395**, 693–698 (1998).
- Stopfer, M., Bhagavan, S., Smith, B. & Laurent, G. Impaired odour discrimination on desynchronization of odour-encoding neural assemblies. *Nature* **390**, 70–74 (1997).
- Lawless, H. T. in *Tasting and Smelling* (eds Beauchamp, G. K. & Bartoshuk, L.) (Academic, San Diego, 1997).
- Kay, S. M. *Modern Spectral Estimation* (Prentice-Hall, Englewood Cliffs, New Jersey, 1988).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank B. Smith for help with the statistical analysis of the data. This work was supported by an NRSA from the NIDCD (M.S.), and the NIDCD and the Alfred P. Sloan Foundation (G.L.).

Correspondence and requests for materials should be addressed to G.L. (e-mail: laurentg@its.caltech.edu).

Phosphorylation of DARPP-32 by Cdk5 modulates dopamine signalling in neurons

James A. Bibb*, Gretchen L. Snyder*, Akinori Nishii, Zhen Yan*, Laurent Meijer†, Allen A. Fienberg*, Li-Huei Tsai§, Young T. Kwon§, Jean-Antoine Girault||, Andrew J. Czernik*, Richard L. Huganir¶, Hugh C. Hemmings Jr#, Angus C. Nairn* & Paul Greengard*

*Laboratory of Molecular and Cellular Neuroscience, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

†Department of Physiology, Kurume University School of Medicine, 67 Asahi-machi Kurume, Fukuoka 830-0011, Japan

‡Centre National de la Recherche Scientifique, Station Biologique, B.P.74, 29682 Roscoff, France

§Department of Pathology, Harvard Medical School, 200 Longwood Avenue, D2-342, Boston, Massachusetts 02115, USA

||INSERM U114, Collège de France, 11 Place Marcelin Bethelot, 75231 Paris Cedex 05, France

¶Department of Neuroscience, Howard Hughes Medical Institute, Johns Hopkins University School of Medicine, 725 N. Wolfe Street, Baltimore, Maryland 21205-2185, USA

#Departments of Anesthesiology and Pharmacology, Weill Medical College of Cornell University, 525 E. 68th Street, New York, New York 10021, USA

The physiological state of the cell is controlled by signal transduction mechanisms which regulate the balance between protein kinase and protein phosphatase activities¹. Here we report that a single protein can, depending on which particular amino-acid residue is phosphorylated, function either as a kinase or phosphatase inhibitor. DARPP-32 (dopamine and cyclic AMP-regulated phospho-protein, relative molecular mass 32,000) is converted into an inhibitor of protein phosphatase 1 when it is phosphorylated by protein kinase A (PKA) at threonine 34 (refs 2, 3). We find that DARPP-32 is converted into an inhibitor of PKA

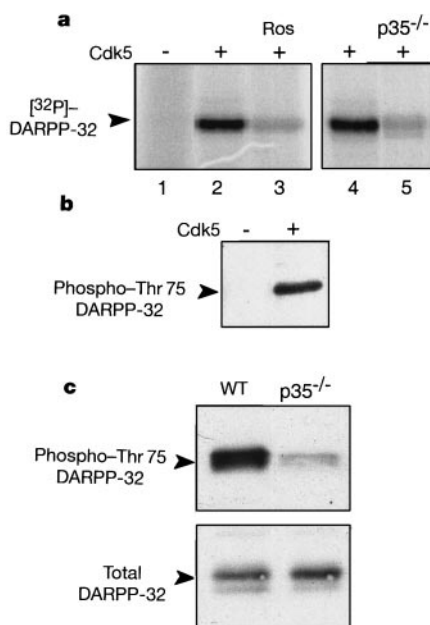


Figure 1 Phosphorylation of DARPP-32 at Thr 75 by Cdk5 *in vitro* and *in vivo*. **a**, Phosphorylation of DARPP-32 by Cdk5 immunoprecipitated from striatum. Roscovitine (Ros) and Cdk5 from p35^{-/-} mice were used as indicated. **b**, Phosphorylation of DARPP-32 at Thr 75 *in vitro* by recombinant Cdk5. **c**, Phosphorylation of DARPP-32 at Thr 75 in intact striatum. Striatal tissue from wild-type (WT) or p35^{-/-} mice were immunoblotted and probed with antibodies against phospho-Thr 75 (top) and total DARPP-32 (bottom).

when phosphorylated at threonine 75 by cyclin-dependent kinase 5 (Cdk5). Cdk5 phosphorylates DARPP-32 *in vitro* and in intact brain cells. Phospho-Thr 75 DARPP-32 inhibits PKA *in vitro* by a competitive mechanism. Decreasing phospho-Thr 75 DARPP-32 in striatal slices, either by a Cdk5-specific inhibitor or by using genetically altered mice, results in increased dopamine-induced phosphorylation of PKA substrates and augmented peak voltage-gated calcium currents. Thus DARPP-32 is a bifunctional signal transduction molecule which, by distinct mechanisms, controls a serine/threonine kinase and a serine/threonine phosphatase.

Numerous first messengers modulate the phosphorylation and dephosphorylation of DARPP-32 at Thr 34 in intact cells, thereby altering activity of protein phosphatase 1 (PP1) and regulating the states of phosphorylation and activities of a variety of downstream physiological effectors³. The biological significance of this pathway has been demonstrated using mice containing a targeted deletion of the DARPP-32 gene (DARPP-32^{-/-}), which exhibit a markedly altered biochemical, electrophysiological and behavioural phenotype⁴.

The amino-acid sequence of DARPP-32 contains consensus phosphorylation sites for proline-directed kinases, including

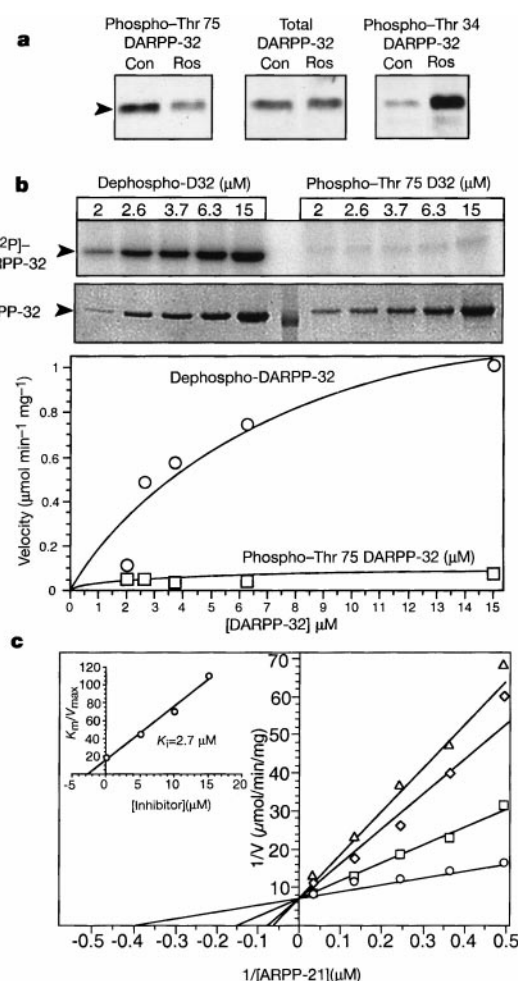


Figure 2 Inhibition of PKA by phospho-Thr 75 DARPP-32. **a**, Effect of roscovitine (Ros) on phospho-Thr 75 and phospho-Thr 34 in mouse striatal slices compared with controls (Con). **b**, Phosphorylation by PKA of dephospho-DARPP-32 (D-32) and phospho-Thr 75 DARPP-32. Top, radiographic image of ³²P-labelled DARPP-32. Middle, Coomassie brilliant blue stain. Bottom, quantification of phosphorylation by PKA of dephospho-DARPP-32 and phospho-Thr 75 DARPP-32. **c**, Lineweaver-Burk analysis of PKA phosphorylation of ARPP-21 in the presence of 0 (circles), 5 (squares), 10 (diamonds) or 15 (triangles) μ M phospho-Thr 75 DARPP-32. Inset, secondary plot from which the K_i value was derived.

Cdk5, a cyclin-dependent kinase family member which, together with its non-cyclin cofactor p35, is present in post-mitotic neurons expressing high levels of DARPP-32 (refs 5, 6). Cdk5, immunoprecipitated from mouse striatal homogenates, phosphorylated purified DARPP-32 *in vitro* at a linear rate over at least 60 min (Fig. 1a; and data not shown). Phosphorylation was greatly reduced by the addition of the specific Cdk5 inhibitor, roscovitine⁷. DARPP-32 was much less efficiently phosphorylated by Cdk5 immunoprecipitated from striatal homogenates of p35^{-/-} mice (Fig. 1a).

Purified DARPP-32 was phosphorylated *in vitro* by recombinant glutathione S-transferase (GST)–Cdk5 (Fig. 1b) and by recombinant Cdk1. Proteolytic digestion, high-performance liquid chromatography purification, phosphopeptide mapping, microsequencing, and MALDI-TOF matrix-assisted laser desorption ionization-time of flight mass spectrometry analysis indicated that the site of phosphorylation by either Cdk5 or Cdk1 was Thr 75. A recombinant protein containing a Thr75Ala mutation was not phosphorylated, indicating that Thr 75 is the only site phosphorylated by Cdk5 and Cdk1. In contrast, mitogen-activated protein kinase, another proline-directed kinase, did not phosphorylate DARPP-32 at a significant rate, and no phosphorylation at Thr 75 could be detected (data not shown).

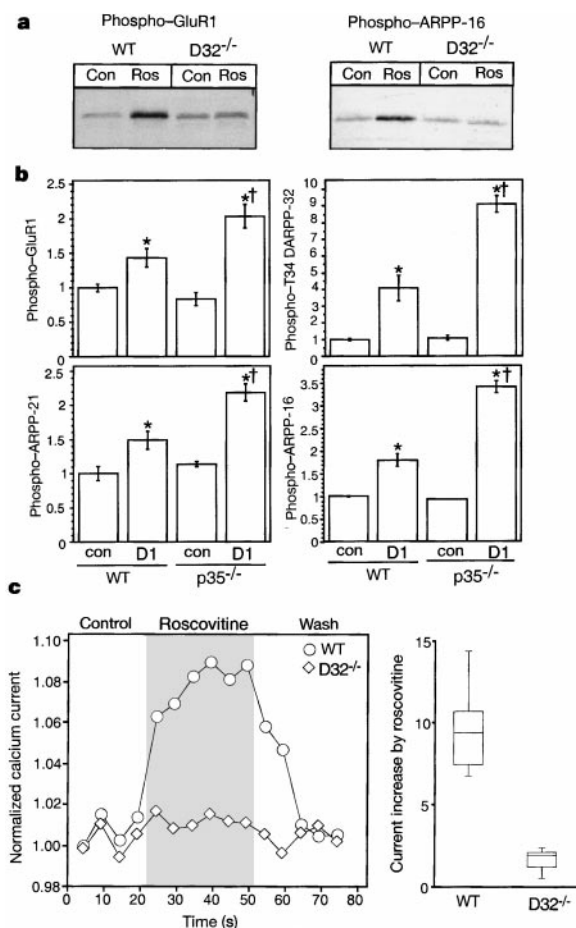


Figure 3 Effect of phospho-Thr 75 DARPP-32 on PKA activity in intact neurons. **a**, Effect of roscovitine on phosphorylation of GluR1 and ARPP-16 in striatal slices from wild-type and DARPP-32^{-/-} mice. **b**, Effect of the selective D1 agonist SKF 81297 (D1) on PKA activity in wild-type and p35^{-/-} mice. Antibodies against the PKA phosphorylation sites of proteins indicated on the left of each panel were used (phospho-GluR1, *n* = 4; phospho-Thr 34 DARPP-32, *n* = 6; phospho-ARPP-21, *n* = 4; phospho-ARPP-16, *n* = 4; asterisk, *P* < 0.05 versus control; cross, *P* < 0.05 versus wild-type D1, by analysis of variance, Newman–Kuel post-hoc test). **c**, Effect of roscovitine on whole-cell Ca²⁺ current in striatal neurons. Left, peak Ca²⁺ current versus time. Right, summary of the roscovitine-induced increase in Ca²⁺ current.

To monitor DARPP-32 phosphorylation by Cdk5 directly *in vitro* and in brain tissue, we generated a phosphorylation state-specific antibody which detected DARPP-32 only in its Thr 75-phosphorylated state (Fig. 1b). DARPP-32 was phosphorylated at Thr 75 in freshly prepared adult mouse striatal tissue (Fig. 1c). Comparison of phospho-Thr 75 and total DARPP-32 standards indicated that the stoichiometry of Thr 75 DARPP-32 phosphorylation was ~0.26 under basal conditions. The level of DARPP-32 in striatal neurons has been estimated to be ~50 μ M (ref. 8). Therefore, the basal concentration of phospho-Thr 75 DARPP-32 was ~13 μ M. The basal level of phospho-Thr 75 was reduced by 75% in p35^{-/-} mouse striatum, whereas the total level of DARPP-32 was unaffected (Fig. 1c).

The role of DARPP-32 phosphorylated at Thr 75 was assessed in striatal slices from adult mice. Inhibition of Cdk5 by 10 μ M roscovitine reduced phospho-Thr 75 DARPP-32 by 73% without affecting the total level of DARPP-32 (Fig. 2a). Remarkably, 10 μ M roscovitine also increased phospho-Thr 34 DARPP-32 10-fold from a basal stoichiometry of 0.01 mol mol⁻¹. Dose–response studies indicated that this concentration of roscovitine caused a maximal increase in phospho-Thr 34 DARPP-32 and a maximal decrease in phospho-Thr 75 DARPP-32. We used a biochemical approach to determine the mechanism underlying this reciprocal relationship.

The ability of PKA to phosphorylate Thr 34 of DARPP-32 *in vitro* was virtually abolished by prior phosphorylation at Thr 75 (Fig. 2b). To test whether this effect was due to phospho-Thr 75 acting as an inhibitor of PKA, the ability of the purified catalytic subunit of PKA to phosphorylate several other well-characterized substrates was assessed in the absence and presence of added phospho-Thr 75 DARPP-32. The presence of 15 μ M phospho-Thr 75 DARPP-32 markedly reduced the rates of phosphorylation by PKA of inhibitor-1 (a DARPP-32 homologue²), ARPP-21 (a PKA substrate enriched in the basal ganglia⁹), synapsin I and histone H1 (data not shown). Kinetic analysis of ARPP-21 phosphorylation in the presence of various concentrations of phospho-Thr 75 DARPP-32 indicated that phospho-Thr 75 DARPP-32 inhibited PKA by a linear competitive single-site mechanism (Fig. 2c). A *K_i* value of 2.7 μ M was derived from a secondary plot of the kinetic data (Fig. 2c, inset).

To determine whether phospho-Thr 75 DARPP-32 inhibits PKA in intact neurons, we assessed phosphorylation of known substrates in mouse striatal slices treated with roscovitine to inhibit Cdk5. Roscovitine increased phosphorylation of the GluR1 subunit of the

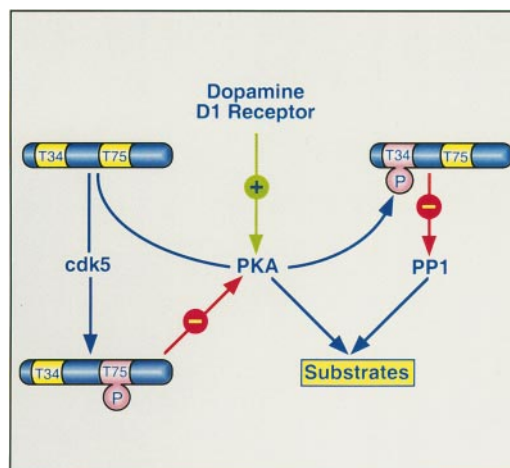


Figure 4 Model illustrating dual effects of phosphorylation of DARPP-32 in regulation of PKA and PP1 activities. Dopamine D1 receptor activation of PKA and phosphorylation of Thr 34 by PKA converts DARPP-32 into an inhibitor of PP1. The activation of PKA and inhibition of PP1 synergistically increase phosphorylation of various substrates³. Conversely, phosphorylation of DARPP-32 at Thr 75 by Cdk5 causes inhibition of PKA and activation of PP1, and synergistically reduces phosphorylation of various substrates.

AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptor 3.2 ± 0.5 -fold ($n = 7$), as assessed using an antibody specific for phospho-Ser845, a PKA phosphorylation site¹⁰ (Fig. 3a). No roscovitine-induced increase in phospho-Ser845 GluR1 was observed in striatal slices from adult DARPP-32^{-/-} mice⁴, providing strong support for the involvement of phospho-Thr75 in this action of roscovitine. Total GluR1 levels were unaffected (data not shown). Similar results (2.5 ± 0.2 -fold, $n = 2$) for wild-type and DARPP-32^{-/-} mice were observed for phosphorylation of ARPP-16 (a PKA substrate enriched in striatum¹¹). These results indicate that regulation by Cdk5 of the phosphorylation of DARPP-32 at Thr75 has a strong effect on PKA activity in intact cells.

Dopamine achieves many of its effects in neurons by activating the D1 class of dopamine receptors, which are positively coupled to the activation of PKA³. To determine whether phospho-Thr75 DARPP-32 modulates the actions of dopamine in intact cells, the ability of a D1 receptor agonist (SKF 81297) to induce phosphorylation of substrates for PKA was compared in wild-type and p35^{-/-} mice. For this purpose, the state of phosphorylation of GluR1, DARPP-32, ARPP-21 and ARPP-16 was determined using phosphorylation state-specific antibodies. SKF 81297 increased the phosphorylation of each of these substrates in the wild-type mice, and did so to an even greater extent in the p35^{-/-} mice (Fig. 3b). No significant difference was observed between wild-type and p35^{-/-} mice in basal levels of phosphorylation of any of these substrates. These data indicate that Cdk5 phosphorylation of DARPP-32 at Thr75 can modulate dopamine signalling in the striatum.

We also assessed the physiological consequences of PKA inhibition by phospho-Thr75 DARPP-32. Voltage-gated Ca²⁺ channels, which are regulated by PKA, are important targets for dopaminergic signalling in the striatum^{12–14}. Therefore, we analysed this class of channel using patch-clamp recordings of dissociated striatal neurons (Fig. 3c). Application of roscovitine enhanced whole-cell Ca²⁺ current in wild-type neurons ($9.7 \pm 1.0\%$, $n = 7$). This effect was virtually abolished in neurons from DARPP-32 knockout mice ($1.6 \pm 0.3\%$, $n = 5$, $P < 0.01$, paired t -test). The dose–response curve for the effect of roscovitine on Ca²⁺ influx was similar to those observed for regulation of phospho-Thr75 and phospho-Thr34 of DARPP-32, with a maximal effect at $10 \mu\text{M}$ (data not shown). These data indicate that phosphorylation of Thr75 of DARPP-32 may be able to modulate the physiological state of striatal neurons.

Protein kinases and phosphatases are localized to their substrates by targeting and scaffolding molecules^{15,16}. In addition, protein kinases and phosphatases associate with each other in signalling modules¹⁷. Our data show that, through phosphorylation of distinct sites, DARPP-32 can regulate both classes of activity (Fig. 4). This dual action appears to be especially important in regulating the efficacy of dopaminergic neurotransmission. Our results also indicate a novel role for Cdk5 in the actions of dopamine. The ability of one protein to regulate both an important protein kinase and an important protein phosphatase represents a new mechanism by which cell signalling pathways may be integrated. □

Methods

Purified recombinant DARPP-32 was phosphorylated by Cdk5 immunoprecipitated from 100 μg of acutely dissected striatal homogenate as described¹⁸. Immunoblotting showed that the immunoprecipitated complex from wild-type mice contained both Cdk5 and its cofactor p35 (antibodies from Santa Cruz Biotechnology, data not shown). Reaction mixtures containing 600 $\mu\text{Ci ml}^{-1}$ [γ -³²P]ATP were incubated for 0–60 min, with 50 μM roscovitine where applicable, followed by addition of gel loading buffer (final concentrations, 1% SDS, 12% glycerol, 25 mM DTT) and boiling for 5 min. Samples were subjected to SDS-PAGE (15% acrylamide), gels were dried, and radiographic images were generated using a PhosphorImager (Molecular Dynamics). DARPP-32 was also phosphorylated by recombinant Cdk5 in reaction mixtures containing 10 μM DARPP-32, 200 μM ATP, 30 mM MOPS (pH 7.2) and 5 mM MgCl₂. Samples were incubated for 60 min with or without partially purified GST–Cdk5 and GST–p25 (active fragment of p35) at 30 °C. Reaction mixtures were processed as above, followed by electrophoretic transfer to nitrocellulose and immunoblotting using a phospho-Thr75 phosphorylation state-specific antibody prepared as described¹⁹.

For kinetic analysis of PKA phosphorylation of dephospho- and phospho–Thr75

DARPP-32, purified protein was phosphorylated to a stoichiometry of 0.93 mol mol⁻¹ using native Cdk1/Cyclin B purified from sea star. Phospho-Thr75 DARPP-32 was precipitated from the reaction mixture with 5% TCA, resuspended in 1 M Tris-HCl (pH 8.8), and dialysed against 10 mM HEPES (pH 7.4). Dephospho- and phospho–Thr75 DARPP-32 were phosphorylated by PKA in the presence of [γ -³²P]ATP as described²⁰. For kinetic analysis of inhibition of PKA phosphorylation of ARPP-21, recombinant ARPP-21 was used and reactions were carried out as described⁹. Phosphorylation was quantified by PhosphorImager analysis.

We prepared striatal slices using standard methodology²¹. Slices from either wild-type or DARPP-32^{-/-} mice were incubated for 1 h in the absence or presence of 10 μM roscovitine. Slices from wild-type and p35^{-/-} mice were incubated for 10 min in the absence or presence of 1 μM SKF 81297. Slices were homogenized by sonication in boiling 1% SDS and 50 mM NaF. Equal amounts of protein were processed as described using 10–20% acrylamide gradient gels and immunoblots were carried out with antibodies against phospho-Ser845 GluR1 (ref. 10), phospho-Thr34 DARPP-32 (ref. 22), phospho–Ser55 ARPP-21 (ref. 23), and phospho–Ser88 ARPP-16.

For electrophysiology, peak voltage-gated Ca²⁺ current was recorded in acutely dissociated striatal neurons from wild-type and DARPP-32^{-/-} mice in response to bath application of roscovitine. Standard whole-cell patch-clamp techniques and data analysis were used as described¹².

Received 12 July; accepted 30 September 1999.

- Hunter, T. Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling. *Cell* **80**, 225–236 (1995).
- Hemmings, H. C. Jr, Greengard, P., Tung, H. Y. L. & Cohen, P. DARPP-32, a dopamine-regulated neuronal phosphoprotein, is a potent inhibitor of protein phosphatase-1. *Nature* **310**, 503–505 (1984).
- Greengard, P., Allen, P. B. & Nairn, A. C. Beyond the dopamine receptor: the DARPP-32/protein phosphatase-1 cascade. *Neuron* **23**, 435–447 (1999).
- Fienberg, A. A. et al. DARPP-32: Regulator of the efficacy of dopaminergic neurotransmission. *Science* **281**, 838–842 (1998).
- Lew, J. et al. A brain-specific activator of cyclin-dependent kinase 5. *Nature* **371**, 423–426 (1994).
- Tsai, L.-H., Delalle, I., Caviness, V. S. Jr, Chae, T. & Harlow, E. p35 is a neural-specific regulatory subunit of cyclin-dependent kinase 5. *Nature* **371**, 419–423 (1994).
- Meijer, L. et al. Biochemical and cellular effects of roscovitine, a potent and selective inhibitor of the cyclin-dependent kinases cdc2, cdk2 and cdk5. *Eur. J. Biochem.* **243**, 527–536 (1997).
- Ouimet, C. C., Langley-Gullion, K. C. & Greengard, P. Quantitative immunocytochemistry of DARPP-32-expressing neurons in the rat caudateputamen. *Brain Res.* **808**, 8–12 (1998).
- Hemmings, H. C. Jr, Girault, J.-A., Williams, K. R., LoPresti, M. B. & Greengard, P. ARPP-21, a cyclic AMP-regulated phosphoprotein ($M_r = 21,000$) enriched in dopamine-innervated brain regions: Amino acid sequence of the site phosphorylated by cyclic AMP in intact cells and kinetic studies of its phosphorylation *in vitro*. *J. Biol. Chem.* **264**, 7726–7733 (1989).
- Roche, K. W., O'Brien, R. J., Mammen, A. L., Bernhardt, J. & Huganir, R. L. Characterization of multiple phosphorylation sites on the AMPA receptor GluR1 subunit. *Neuron* **16**, 1179–1188 (1996).
- Horiuchi, A., Williams, K. R., Kurihara, T., Nairn, A. C. & Greengard, P. Purification and cDNA cloning of ARPP-16, a cAMP-regulated phosphoprotein enriched in basal ganglia, and of a related phosphoprotein, ARPP-19. *J. Biol. Chem.* **265**, 9476–9484 (1990).
- Surmeier, D. J., Bargas, J., Hemmings, H. C. Jr, Nairn, A. C. & Greengard, P. Modulation of calcium currents by a D1 dopaminergic protein kinase/phosphatase cascade in rat neostriatal neurons. *Neuron* **14**, 385–397 (1995).
- Gray, P. C., Scott, J. D. & Caterall, W. A. Regulation of ion channels by cAMP-dependent protein kinase and A-kinase anchoring proteins. *Curr. Opin. Neurobiol.* **8**, 330–334 (1998).
- Nicola, S. M., Surmeier, J. D. & Malenka, R. C. Dopaminergic modulation of neuronal excitability in the striatum and nucleus accumbens. *Annu. Rev. Neurosci.* (in the press).
- Pawson, T. & Scott, J. D. Signaling through scaffold, anchoring, and adaptor proteins. *Science* **278**, 2075–2080 (1997).
- Klauck, T. M. et al. Coordination of three signaling enzymes by AKAP79, a mammalian scaffold protein. *Science* **271**, 1589–1592 (1996).
- Westphal, R. S., Anderson, K. A., Means, A. R. & Wadzinski, B. E. A signaling complex of Ca²⁺-calmodulin-dependent protein kinase IV and protein phosphatase 2A. *Science* **280**, 1258–1261 (1998).
- Tsai, L.-H., Takahashi, T., Caviness, V. S. Jr & Harlow, E. Activity and expression pattern of cyclin-dependent kinase 5 in the embryonic mouse nervous system. *Development* **119**, 1029–1040 (1993).
- Czernik, A. J., Mathers, J., Tsou, K., Greengard, P. & Mische, S. M. In *Regulatory Protein Modification, Neurochemicals* vol. 30 (ed. Hemmings, H. C. Jr) 219–246 (Humana, Totowa, New Jersey, 1997).
- Girault, J.-A., Hemmings, H. C. Jr, Williams, K. R., Nairn, A. C. & Greengard, P. Phosphorylation of DARPP-32, a dopamine- and cAMP-regulated phosphoprotein, by casein kinase II. *J. Biol. Chem.* **264**, 21748–21759 (1989).
- Nishi, A., Snyder, G. L. & Greengard, P. Bidirectional regulation of DARPP-32 phosphorylation by dopamine. *J. Neurosci.* **17**, 8147–8155 (1997).
- Snyder, G. L. et al. Phosphorylation of DARPP-32 and protein phosphatase inhibitor-1 in rat choroid plexus: Regulation by factors other than dopamine. *J. Neurosci.* **12**, 3071–3083 (1992).
- Caporaso, G. et al. Dopamine and drugs of abuse modulate phosphorylation of ARPP-21, a cyclic AMP-regulated phosphoprotein enriched in the neostriatum. *Neuropharmacology* (in press).

Acknowledgements

We thank A. Horiuchi for construction of the Thr75A mutant plasmid; S. Rakhilin for recombinant ARPP-21; I. Dulubova for phospho-Ser88 ARPP-16 antibody; J. Wang for recombinant Cdk5/p25; J. Volker for help in providing p35^{-/-} mice; and the Rockefeller University Protein/DNA Technology Center. This work was supported by a National Research Service Award (J.B.) and funding from the National Institute of Mental Health and the National Institute of Drug Abuse (P.G.).

Correspondence and requests for materials should be addressed to P.G. (e-mail: greengd@rockefeller.edu).

Induction of autophagy and inhibition of tumorigenesis by *beclin 1*

Xiao Huan Liang*, Saadiya Jackson*, Matthew Seaman†, Kristy Brown‡, Bettina Kempkes§, Hanina Hibshoosh‡ & Beth Levine*

* Department of Medicine and ‡ Department of Pathology, Columbia University College of Physicians & Surgeons, New York, New York 10032, USA

† Department of Clinical Biochemistry, Cambridge Institute for Medical Research, Addenbrookes Hospital, Cambridge, CB2 2XY, UK

§ Institute for Clinical Molecular Biology, GSF-National Research Center for Environment and Health, 81377 Munich, Germany

The process of autophagy, or bulk degradation of cellular proteins through an autophagosomal-lysosomal pathway¹, is important in normal growth control and may be defective in tumour cells². However, little is known about the genetic mediators of autophagy in mammalian cells or their role in tumour development. The mammalian gene encoding Beclin 1 (ref. 3), a novel Bcl-2-interacting, coiled-coil protein, has structural similarity to the yeast autophagy gene, *apg6/vps30* (refs 4, 5), and is mono-allelically deleted in 40–75% of sporadic human breast cancers and ovarian cancers⁶. Here we show, using gene-transfer techniques, that *beclin 1* promotes autophagy in autophagy-defective yeast with a targeted disruption of *apg6/vps30*, and in human MCF7 breast carcinoma cells. The autophagy-promoting activity of *beclin 1* in MCF7 cells is associated with inhibition of MCF7 cellular proliferation, *in vitro* clonogenicity and tumorigenesis in nude mice. Furthermore, endogenous Beclin 1 protein expression is frequently low in human breast epithelial carcinoma cell lines and

tissue, but is expressed ubiquitously at high levels in normal breast epithelia. Thus, *beclin 1* is a mammalian autophagy gene that can inhibit tumorigenesis and is expressed at decreased levels in human breast carcinoma. These findings suggest that decreased expression of autophagy proteins may contribute to the development or progression of breast and other human malignancies.

Fourteen genes have been identified in *Saccharomyces cerevisiae* that are required for yeast autophagy⁷. Yeast that are lacking one of these genes, *apg6/vps30*, are defective in both their ability to undergo nitrogen deprivation-induced autophagy⁴ and their ability to properly sort selective vacuolar proteins⁵. Human Beclin 1 shares 24.4% amino-acid identity (and 39.1% conservation) with Apg6/Vps30p. To determine whether Beclin 1 is a functional homologue of Apg6/Vps30, we investigated whether *beclin 1* expression could restore autophagic activity and vacuolar protein sorting in *apg6/vps30*-disrupted yeast. Four hours after transfer to nitrogen-starvation media, we compared the percentage of cells with autophagic bodies within the yeast vacuole in isogenic wild-type yeast (strain SEY620), *apg6/vps30*-disrupted yeast (strain JCY3000), JCY3000 yeast transformed with *apg6/vps30*, JCY3000 yeast transformed with *beclin 1* and JCY3000 yeast transformed with empty vector (Fig. 1). In nitrogen-deprivation conditions, wild-type yeast accumulate autophagic bodies within their central vacuole (the mammalian equivalent of lysosomes) and detection of autophagic bodies is facilitated by phenylmethylsulphonyl fluoride (PMSF), a protease inhibitor that blocks their degradation⁸. In the presence of PMSF, we observed numerous cells with autophagic bodies within the vacuole in wild-type SEY6210 yeast, and in *apg6/vps30* and *beclin-1*-transformed JCY3000 yeast (Fig. 1b, c). In contrast, significantly fewer cells were seen with autophagic bodies within the vacuole among nontransformed JCY3000 yeast or JCY3000 yeast transformed with empty vector ($P < 0.0001$, analysis of variance). These

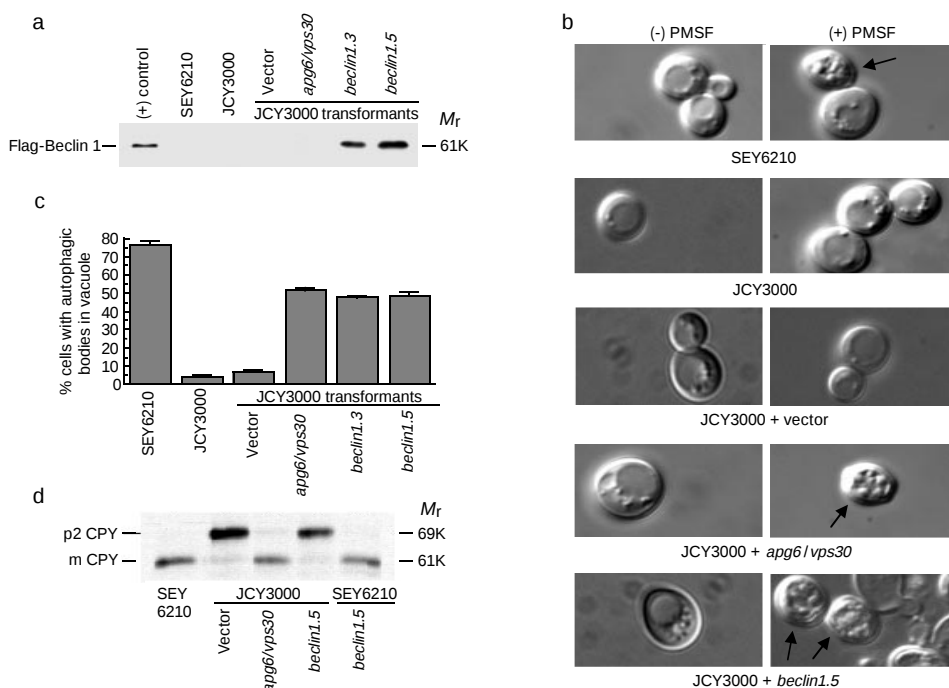


Figure 1 Complementation by *beclin 1* of the yeast autophagy but not vacuolar protein sorting function of *apg6/vps30*. **a**, Western blot analysis of yeast cell lysates with polyclonal anti-flag antibody to detect flag-tagged human Beclin 1 expression. Two-hundred micrograms of protein per lane was loaded. SEY6210, wild-type yeast; JCY3000, SEY6210 disrupted in *apg6/vps30*; (+) control, lysates of mammalian cells transfected with flag-*beclin 1*. **b**, Autophagic body formation in yeast deprived of nitrogen in the presence (+) and absence (-) of 1 mM PMSF. Arrows denote cells that would be scored as positive in the experiment shown in **c**. **c**, Quantitative effects of *apg6/vps30* and

beclin 1 transformation on autophagic body formation in *apg6/vps30*-disrupted yeast in the presence of PMSF. Cells with one or more autophagic bodies within the vacuole were scored as positive (see arrows in **b**). A minimum of 100 cells was counted for each sample. Results represent the mean ($\pm$ s.e.m.) percentage of cells with autophagic bodies within the vacuole for triplicate samples. Similar results were obtained in five independent experiments. **d**, Sorting of vacuolar protein carboxypeptidase Y (CPY). p2CPY, precursor form; mCPY, sorted mature form. M_r , relative molecular mass.

results indicate that *beclin 1* restores autophagy induced by nitrogen deprivation in autophagy-defective *apg6/vps30*-disrupted yeast. However, unlike JCY3000 yeast transformed with *apg6/vps30*, JCY3000 yeast transformed with *beclin 1* were unable to properly sort and mature the vacuolar protein, carboxypeptidase Y (CPY) (Fig. 1d). Thus, *beclin 1* complements the autophagy but not the vacuolar protein sorting function of *apg6/vps30* in *apg6/vps30*-disrupted yeast.

The *beclin 1* gene maps to a tumour susceptibility locus on human chromosome 17q21 (refs 9, 10) that is mono-allelically deleted in up to 40–75% of cases of sporadic ovarian and breast carcinomas⁶. This observation, coupled with the homology between *beclin 1* and *apg6/vps30*, led us to propose that *beclin 1* is a mammalian autophagy gene that negatively regulates tumorigenesis. Therefore, we investigated the effects of *beclin 1* gene transfer on the autophagic activity (Fig. 2) and growth properties (Fig. 3) of MCF7 human breast carcinoma cells. MCF7 cells were originally derived from a patient with 17q21 loss of heterozygosity¹¹ and do not express detectable levels of endogenous Beclin 1 (Fig. 4a). To permit conditional expression of a potential antiproliferative gene, we stably transfected MCF7 cells with a tetracycline-repressible vector, pTC (ref. 12), containing flag-epitope-tagged full-length human Beclin 1 (referred to as MCF7.*beclin1* clones) (Fig. 3a), flag epitope-tagged *beclin 1* containing a premature stop codon at nucleotide position 270 (MCF7.*beclin1stop* clones) (Fig. 3b) or no insert (MCF7.control clones).

Three MCF7.control clones, three MCF7.*beclin1* clones and one MCF7.*beclin1stop* clone, which were either maintained in normal growth conditions or subjected to serum and amino-acid depriva-

tion for 4 hours were analysed by electron microscopy (Fig. 2). In normal cultured cells, serum and amino-acid deprivation is a potent inducer of autophagy¹³, which can be recognized at the ultrastructural level as double-membrane vacuolar structures containing visible cytoplasmic contents^{14–16}. We found that baseline numbers of autophagic vacuoles per cell were significantly higher in MCF7.*beclin1* compared with MCF7.control or MCF7.*beclin1stop* clones ($P < 0.0001$, analysis of variance; see Fig. 2a, c, g). Furthermore, serum and amino-acid deprivation did not increase the mean number of autophagic vacuoles per cell in MCF7.control or MCF7.*beclin1stop* clones (Fig. 2b, g), but did induce a significant increase in MCF7.*beclin1* clones ($P < 0.0001$, *t*-test; Fig. 2d, g). Although autophagic vacuoles were found in the cytoplasm, the nuclei of MCF7.*beclin1* clones appeared normal, without evidence of apoptotic morphologic changes (Fig. 2c, d; data not shown). The serum and amino-acid deprivation-induced increase in autophagic vacuole formation in MCF7.*beclin1* clones was inhibited by pre-treatment with 3-methyladenine (3-MA), a nucleotide derivative that inhibits the earliest stages of autophagosome formation¹⁷ (Fig. 2g). These data indicate that MCF7 human breast carcinoma cells lack Beclin 1 expression and are unable to undergo nutrient deprivation-induced autophagy. In contrast, enforced *beclin 1* expression promotes nutrient deprivation-induced autophagy in MCF7 cells.

To confirm these morphologic observations using a biochemical approach, we measured the rates of intracellular degradation of long-lived proteins following nutrient deprivation of MCF7.*beclin1*, MCF7.control and MCF7.*beclin1stop* clones (Fig. 2h). In response to nutrient deprivation or high density culture conditions,

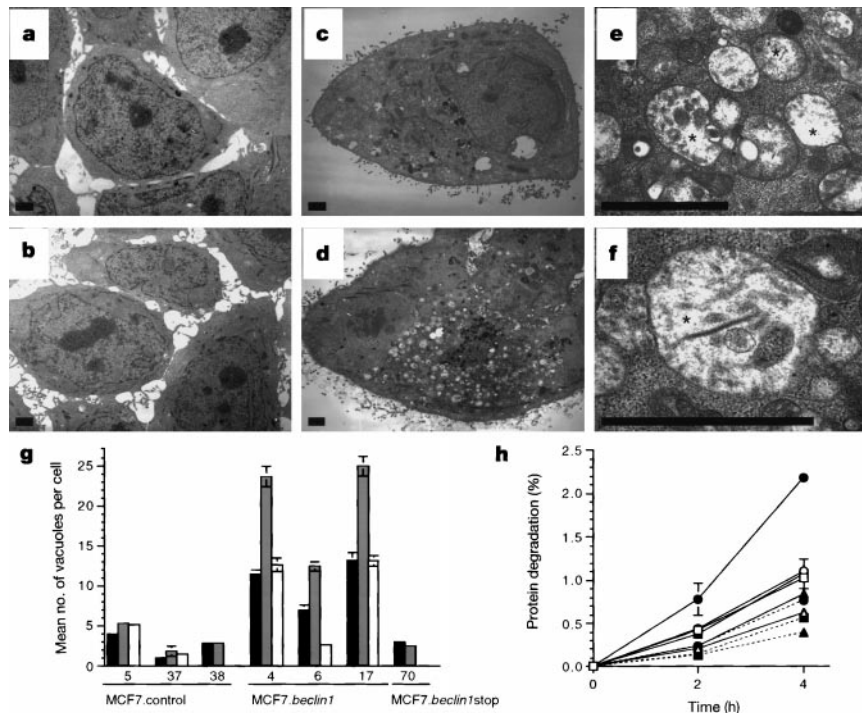


Figure 2 Promotion of autophagy by enforced *beclin 1* expression in human MCF7 breast carcinoma cells. **a–f**, Electron micrographs of MCF7.control (clone 38) (**a, b**) and MCF7.*beclin1* (clone 17) cells (**c–f**) grown in nutrient-rich media (**a, c**) or subjected to 4 h of serum and amino-acid deprivation (**b, d–f**). Asterisks in **e** and **f** denote autophagic vacuoles that would be counted in the experiment shown in **g**. Autophagic vacuoles were defined as double-membrane vacuolar structures containing recognizable cytoplasmic contents. Scale bars, 1 μ m. **g**, Quantitative effects of *beclin 1* on basal and nutrient-deprivation-induced autophagy of MCF7 cells. Bars indicate mean ($\pm$ s.e.m.) number of autophagic vacuoles per cell for cells growing in normal media (black), for cells subjected to 4 h of serum and amino-acid deprivation (grey), and for cells pre-treated for 30 min with

10 mM 3-MA and subjected to 4 h of serum and amino-acid deprivation in the presence of 3-MA (open). Mean was determined by counting the total number of autophagic vacuoles in each cell for 100 cells per clone per treatment. **h**, Comparison of the rates of degradation of long-lived proteins in MCF7.control cells (squares, clone 38), MCF7.*beclin1* cells (circles, clone 17) and MCF7.*beclin1stop* cells (triangles, clone 70), in EBSS + 10% serum and complete amino acids (open symbols, solid lines) in EBSS alone (closed symbols, solid lines), and in EBSS + 10 mM 3-MA (closed symbols, dotted lines). Results are mean ($\pm$ s.e.m.) of triplicate wells. Similar results were obtained in three independent experiments. Results shown with these clones are representative of results obtained with other clones analysed by electron microscopy in **g**.

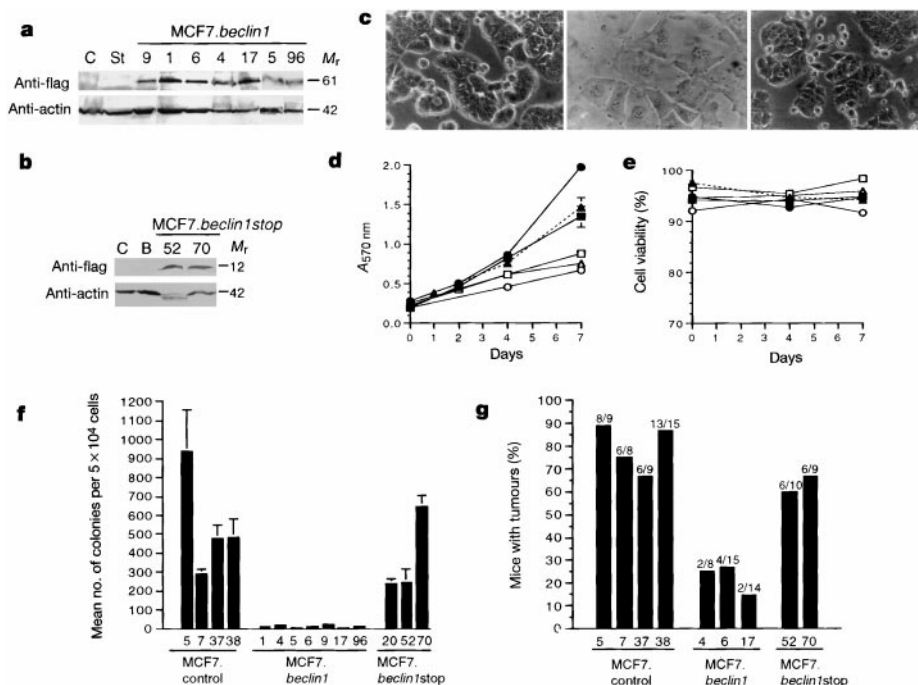


Figure 3 Effects of enforced *beclin 1* expression on the growth properties and tumorigenicity of MCF7 cells. **a**, Western blot analysis of flag-tagged *beclin-1*-transfected MCF7 clones. C, MCF7.control cells (clone 38); St, MCF7.*beclin1stop* cells (clone 70). *M_r*, relative molecular mass. **b**, Western blot analysis of flag-tagged *beclin1stop*-transfected MCF7 clones. C, MCF7.control cells (clone 38); B, MCF7.*beclin1* cells (clone 1). **c**, Photomicrographs of MCF7.control cells (left panel, clone 38), MCF7.*beclin1* cells (middle panel, clone 17) and MCF7.*beclin1stop* cells (right panel, clone 70) 48 h after seeding at similar densities. **d**, Proliferation of MCF7.control cells (clones 37 and 38; closed squares and circles, respectively, solid lines), MCF7.*beclin1* cells (clones 1, 6, 17;

open squares, circles, triangles, respectively, solid lines) and MCF7.*beclin1stop* cells (clone 70, closed triangles, dotted line). Results represent mean Absorbance (A) ($\pm$ s.e.m.) for six wells; similar results were obtained in three independent experiments. **e**, Cell viability of MCF7 clones determined by trypan blue staining. Symbols for each clone are as in **d**. **f**, Clonogenicity in semisolid medium (soft agar) of MCF7 clones. Results are mean ($\pm$ s.e.m.) for pooled triplicate wells from 3–4 independent experiments. **g**, Tumour formation in NCR nude mice injected subcutaneously with MCF7 clones. Numbers on top of bar indicate number of autopsy-confirmed tumours at 8 weeks per number of mice injected with each clone.

increased autophagy results in enhanced proteolysis of long-lived proteins, and this response is often impaired in transformed cells^{2,18–23}. We found that MCF7.*beclin1* clones, but not MCF7.control or MCF7.*beclin1stop* clones, had a significant increase in rate of intracellular proteolysis following nutrient deprivation ($P < 0.0001$, *t*-test), and that this increase was completely blocked by the autophagy inhibitor 3-MA. Thus, during nutrient deprivation the effects of *beclin 1* on autophagic vacuole formation in MCF7 cells are associated with enhanced degradation of long-lived cellular proteins.

To evaluate the effect of *beclin 1* on MCF7 growth properties, we compared the morphology, cellular proliferation rates, *in vitro* clonogenicity and *in vivo* tumorigenicity of MCF7.*beclin1*, MCF7.control and MCF7.*beclin1stop* clones. MCF7.*beclin1* clones displayed several morphologic characteristics consistent with a less malignant phenotype, including flatter appearance, larger size, firmer attachment to tissue culture plate and increased contact inhibition (Figs 2a, c, 3c). Consistent with the autophagy-promoting effects of *beclin 1*, MCF7.*beclin1* clones proliferated more slowly than MCF7.control or MCF7.*beclin1stop* clones (Fig. 3d). This lower rate of proliferation was not explained by increased cell death, as cell viability determined by trypan blue staining was similar in MCF7.*beclin1*, MCF7.*beclin1stop* and MCF7.control cells (Fig. 3e). All MCF7.*beclin1* clones were also severely impaired in their clonogenicity *in vitro*, when compared with MCF7.control and MCF7.*beclin1stop* clones which formed colonies in soft agar with high efficiency ($P < 0.0001$, chi-squared test) (Fig. 3f). Moreover, the incidence of tumour formation of three MCF7.*beclin1* clones that were injected into nude mice (14–27%) was significantly lower than that of four MCF7.control clones (60–89%) and two MCF7.*beclin1stop* clones (60–67%) ($P < 0.0001$; chi-squared test)

(Fig. 3g). Together, these data show that *beclin 1* can function as a negative regulator of mammary cell growth and tumorigenesis.

The frequent allelic deletions of *beclin 1* in human breast carcinoma, coupled with its autophagy-promoting and growth inhibitory effects, suggest that functional inactivation of Beclin 1 may be important in breast cancer development or progression. A previous study did not identify any coding region mutations in *beclin 1* genomic DNA or abnormalities in *beclin 1* messenger RNA expression in human breast carcinoma cell lines⁶. To evaluate whether levels of Beclin 1 protein expression are decreased in human breast carcinoma, we analysed by western blot human breast carcinoma-derived cell lines and matched normal breast and breast carcinoma tissue from patients with invasive sporadic breast carcinoma. Out of 11 human breast carcinoma cell lines examined, only three had detectable Beclin 1 protein expression (Fig. 4a). Out of 17 pairs of equivalent volumes of matched normal breast and breast carcinoma tissue, 15 samples had higher levels of Beclin 1 in normal than in tumour breast tissue (Fig. 4b). In contrast, 16 tumour samples had higher expression of control proteins (for example, actin, cytokeratin) than in the matched normal tissue samples. Therefore, despite higher overall levels of protein in breast carcinomas (which consist primarily of malignant epithelial cells) than in normal breast tissue (which consists primarily of fat cells), specific expression of Beclin 1 protein is usually decreased in breast carcinomas.

To determine whether the decreased levels of Beclin 1 protein expression in breast carcinoma samples reflected a loss of expression specifically in carcinoma cells rather than in stroma or non-neoplastic epithelium, we immunohistochemically stained tissue sections. Adjacent paraffin-embedded sections from matched normal breast and breast carcinoma samples, including 10 of the

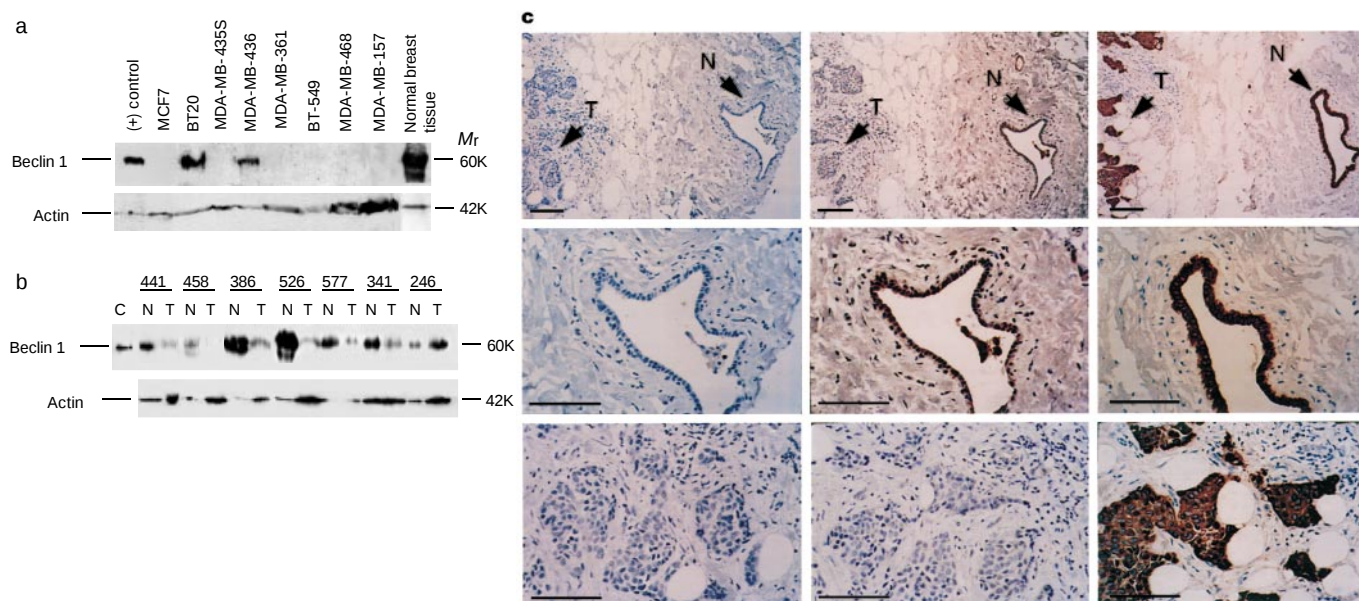


Figure 4 Decrease in Beclin 1 protein expression in human breast carcinoma cell lines and tissue. **a**, Western blot analysis of human breast carcinoma cell lines. Results are from 8 representative cell lines out of a total of 11 analysed. **b**, Western blot analysis of matched normal breast and breast carcinoma tissue from patients with sporadic invasive breast carcinoma. Results are from 7 representative cases out of a total of 17 analysed. N, normal breast tissue; T, tumour tissue; *M_r*, relative molecular mass. **c**, Immunohisto-

chemical analysis of paraffin-embedded sections of breast tissue from patient 441 shown in **b**. Adjacent sections are stained with pre-immune 843 rabbit serum (left column), 843 anti-Beclin 1 serum (middle column) and anti-cytokeratin (right column). Top row shows low power view of region of breast containing tumour (T) and normal mammary epithelial duct (N). Middle row shows higher power view of normal mammary epithelial duct. Bottom row shows higher power view of tumour. Scale bars, 1 mm.

samples analysed by western blot and 22 additional samples, were stained with anti-Beclin 1 and anti-cytokeratin. We found that in all 32 cases, there was strong Beclin 1 immunoreactivity in all normal breast epithelial cells and strong cytokeratin immunoreactivity in all neoplastic as well as normal breast epithelial cells (see Fig. 4c). (Beclin 1 immunoreactivity was also seen in the media of vessels and in some inflammatory cells.) However, in 18 out of 32 cases, there was a significant decrease in the amount of Beclin 1 protein expression in breast carcinoma cells compared with normal breast lobular or ductal epithelial cells (Fig. 4c; Table 1). Thus, *beclin 1* is expressed ubiquitously in normal breast epithelial cells, but its expression is commonly decreased in breast carcinoma cells.

In summary, our data indicate that, to our knowledge, *beclin 1* is the first identified mammalian gene with a role in mediating autophagy. Earlier studies showed that signalling through the S6 kinase pathway exerts inhibitory effects on mammalian autophagy²⁴, but no genes were identified that induced autophagy in mammalian cells. Our findings, that enforced expression of an autophagy gene not only promotes nutrient deprivation-induced autophagy in human breast carcinoma cells but also inhibits their tumour-forming potential, indicate that autophagy may be a fundamental mechanism for preventing the deregulated growth of tumour cells. Further studies are necessary to determine whether induction of autophagy by *beclin 1* is, in fact, the mechanism by which it inhibits tumorigenesis. Nonetheless, the frequent mono-allelic deletions and decreased expression of *beclin 1* in human

breast carcinoma suggest that specific molecular alterations in autophagy pathways may contribute to tumorigenesis. □

Methods

Yeast strains, media and genetic methods

The *S. cerevisiae* strains used for cloning, immunochemical analysis and autophagy assays were SEY6210 (*MAT α leu2-3, 112 ura3-52 his Δ 200 trp1- Δ 901 lys2-801 suc2- Δ 9*) and JCY3000 (*SEY6210 Δ vps30::HIS3*)⁵. JCY3000 yeast were transformed with pRS424.vps30 (ref. 5) as described²⁵, and flag-*beclin 1* was substituted for the open reading frame of *apg6/vps30* in pRS424.vps30 using recombination-mediated PCR-directed plasmid construction *in vivo*²⁶ to generate the two clones JCY3000+*beclin1.3* and JCY3000+*beclin1.5*. Yeast were grown in YEPD, in synthetic complete medium lacking tryptophan for transformation selection (SC – trp), or in SD(–N) for starvation experiments, consisting of 2% glucose with 1.7 g l^{–1} of yeast nitrogen base without amino acids and ammonium sulfate.

Mammalian cell lines and transfections

Human breast carcinoma cell lines were obtained from ATCC (American Type Culture Collection) and maintained according to ATCC instructions. MCF7 cells were transfected using liposome-mediated transfer with the plasmid, pTC (ref. 12), containing flag epitope-tagged human *beclin 1* cloned into the *NheI* site or no insert. Stable transfectants were selected for with 300 μ g ml^{–1} hygromycin and maintained in 2 μ g ml^{–1} tetracycline; tetracycline was withdrawn 5 days before all experiments.

Immunochemical procedures

Yeast and mammalian cell lysates were separated by SDS-PAGE and immunoblotted with anti-flag monoclonal (M2, VWR; 20 μ g ml^{–1}), polyclonal anti-flag antibody (Zymed, 1:200 dilution), polyclonal anti-Beclin 1 antibody 843 (ref. 3) (1:200 dilution) or anti-actin antibody (Boehringer Mannheim, 1:400 dilution). For immunoperoxidase staining to detect endogenous Beclin 1 expression, we used 843 (1:200) and the ABC method (Vector Laboratories). Control staining was performed with pre-immune 843 serum (1:200) and anti-cytokeratin (1:100, Boehringer Mannheim).

Autophagy analysis

Yeast strains were grown overnight in YEPD or liquid SC – trp, transferred to SD(–N) at a concentration of 2×10^7 cells ml^{–1}, incubated for 4 h in the presence or absence of 1 mM PMSE, visualized using DIC optics with a Plan-Apochromat 100 $\times$ 1.4 NA objective, and imaged with a cooled CCD camera using IP Lab software. MCF7 cells were grown in normal media or serum and amino-acid free media, fixed with 2.5% glutaraldehyde, postfixed with 1% OsO₄ and embedded in LX-112 (Ladd Research Industries, Inc.) and Embed-812 (E.M.S.). Thin sections were cut on MT-700 RMC, stained with uranyl acetate and lead citrate, and examined by transmission electron microscopy using a JEOL JEM-1200 EXII.

Table 1 Level of Beclin 1 immunoreactivity in different regions of breast tissue samples from patients with invasive breast carcinoma

Beclin 1 immunoreactivity	Blood vessels	Luminal ducts	No. of cases (%)	
			Luminal lobules	Tumour*
0 (absent)	0 (0.0)	0 (0.0)	0 (0.0)	9 (28.1)
1+ (mild)	1 (3.1)	11 (34.4)	9 (28.1)	16 (50.0)
2+ (moderate)	12 (37.5)	16 (50.0)	18 (56.3)	5 (15.6)
3+ (strong)	19 (59.4)	5 (15.6)	5 (15.6)	2 (6.3)

* $P < 0.0001$ for the comparison of tumour with luminal ducts/luminal lobules; chi-squared test.

Vacuolar protein sorting

Sorting of the vacuolar protein, CPY, was measured as described⁵, with the exception that yeast cells were not separated into intracellular and extracellular fractions.

Measurement of degradation of long-lived proteins

Long-lived proteins were labelled by incubating cells for 72 h in media containing reduced concentrations of unlabelled leucine (0.065 mM) and [³H] leucine (1 μ Ci ml⁻¹). Cells were washed, incubated for 24 h in medium containing 2 mM unlabelled leucine to allow degradation of short-lived proteins and removal of unincorporated labelled leucine, and then subjected to 4-h incubation in normal media or deprivation conditions. Excess unlabelled leucine (2 mM) was added to prevent reincorporation of labelled degraded proteins. At 0, 2 and 4 h, aliquots of the media were measured for acid-soluble radioactivity. At the end of the incubation, the acid-precipitable radioactivity of the cell monolayers was measured and the percentage of long-lived protein degradation was calculated using a described formula²¹.

Proliferation, cell viability, clonogenicity and tumorigenicity assays

MCF7 clones were seeded in 96-well plates at a density of 5×10^3 cells per well, and cell proliferation was measured at serial time points using the MTT Cell Proliferation Kit (Boehringer Mannheim). At identical time points, triplicate wells were subjected to trypan blue staining to determine cell viability. MCF7 clones were plated in triplicate at a density of 5×10^4 cells per 35-mm well in semisolid medium (soft agar) as described²² and colonies were counted at 21 days. Five week-old NCR nude mice (Taconic Farms) were implanted with slow release oestrogen pellets (1.7 mg per 60 days) and injected subcutaneously with 5×10^6 tumor cells. Animals were necropsied 8 weeks after injection for histologic confirmation of tumour.

Received 3 June; accepted 29 September 1999.

- Dunn, W. A. J. Autophagy and related mechanisms of lysosomal-mediated protein degradation. *Trends Cell Biol.* **4**, 139–143 (1994).
- Kisen, G. O. et al. Reduced autophagic activity in primary rat hepatocellular carcinoma and ascites hepatoma cells. *Carcinogenesis* **14**, 2501–2505 (1993).
- Liang, X. H. et al. Protection against fatal Sindbis virus encephalitis by Beclin, a novel Bcl-2-interacting protein. *J. Virol.* **72**, 8586–8596 (1998).
- Kametaka, S., Okano, T., Ohsumi, M. & Ohsumi, Y. Apg14p and Apg6/Vps30p form a protein complex essential for autophagy in the yeast, *Saccharomyces cerevisiae*. *J. Biol. Chem.* **273**, 22284–22291 (1998).
- Seaman, M. N. J., Marcussen, E. G., Cereghino, J. L. & Emr, S. D. Endosome to golgi retrieval of the vacuolar protein sorting receptor, Vps10p, requires the function of the VPS29, VPS30, and VPS35 gene products. *J. Cell. Biol.* **137**, 79–92 (1997).
- Aita, V. M. et al. Cloning and genomic structure of beclin 1, a candidate tumor suppressor gene on chromosome 17q21. *Genomics* **59**, 59–65 (1999).
- Tsukada, M. & Ohsumi, Y. Isolation and characterization of autophagy-defective mutants of *Saccharomyces cerevisiae*. *FEBS* **333**, 169–174 (1993).
- Kametaka, S., Matsuura, A., Wada, Y. & Ohsumi, Y. Structural and functional analyses of APG5, a gene involved in autophagy in yeast. *Gene* **178**, 139–143 (1996).
- Friedman, L. S. et al. The search for BRCA1. *Cancer Res.* **54**, 6374–6382 (1994).
- Rommens, J. M. et al. Generation of a transcription map at the HSD17B locus centromeric to BRCA1 at 17q21. *Genomics* **28**, 530–542 (1995).
- Holt, J. T. et al. Growth retardation and tumour inhibition by BRCA1. *Nature Genet.* **12**, 298–302 (1996).
- Wu, K.-J., Polack, A. & Dalla-Favera, R. Coordinated regulation of iron-controlling genes, H-ferritin and IRP2, by c-MYC. *Science* **283**, 676–679 (1999).
- Mortimore, G. E. & Poso, A. R. Amino acid control of intracellular protein degradation. *Methods Enzymol.* **166**, 461–476 (1988).
- Dunn, W. A. Studies on the mechanisms of autophagy; formation of the autophagic vacuole. *J. Cell. Biol.* **110**, 1923–1933 (1990).
- Dunn, W. A. Studies on the mechanisms of autophagy; maturation of the autophagic vacuole. *J. Cell. Biol.* **110**, 1935–1945 (1990).
- Mitchener, J. S., Shelburne, J. D., Bradford, W. D. & Hawkins, H. K. Cellular autophagocytosis induced by deprivation of serum and amino acids in HeLa cells. *Am. J. Pathol.* **83**, 485–498 (1976).
- Seglen, P. O. & Gordon, P. B. 3-methyladenine: specific inhibitor of autophagic/lysosomal protein degradation in isolated rat hepatocytes. *Proc. Natl Acad. Sci. USA* **79**, 1889–1892 (1982).
- Bradley, M. O. Regulation of protein degradation in normal and transformed human cells. *J. Biol. Chem.* **252**, 5310–5315 (1977).
- Lee, H.-K., Jones, R. T., Myers, R. A. M. & Marzella, L. Regulation of protein degradation in normal and transformed human bronchial epithelial cells in culture. *Arch. Biochem. Biophys.* **296**, 271–278 (1992).
- Gunn, J. M., Clark, M. G., Knowles, S. E., Hopgood, M. F. & Ballard, F. J. Reduced rates of proteolysis in transformed cells. *Nature* **266**, 58–60 (1977).
- Gronostajski, R. M. & Pardee, A. B. Protein degradation in 3T3 and tumorigenic transformed 3T3 cells. *J. Cell. Physiol.* **119**, 127–132 (1984).
- Knecht, E., Hernandez-Yago, J. & Grisolia, S. Regulation of lysosomal autophagy in transformed and non-transformed mouse fibroblasts under several growth conditions. *Exp. Cell. Res.* **154**, 224–232 (1984).
- Otsuka, H. & Moskowitz, M. Differences in the rates of protein degradation in untransformed and transformed cell lines. *Exp. Cell. Res.* **112**, 127–135 (1978).
- Blommaert, E. F. C., Luiken, J. J. F. P., Blommaert, P. J. E., vanWoerkom, G. M. & Meijer, A. J. Phosphorylation of ribosomal protein S6 is inhibitory for autophagy in isolated rat hepatocytes. *J. Biol. Chem.* **270**, 2320–2326 (1995).
- Eble, R. A simple and efficient procedure for transformation of yeasts. *Biotechniques* **13**, 18–20 (1992).

- Oldenburg, K. R., Vo, K. T., Michaelis, S. & Paddon, C. Recombination-mediated PCR-directed plasmid construction in vivo in yeast. *Nucleic Acids Res.* **25**, 451–452 (1997).
- Jiang, W. & Al, E. Overexpression of cyclin D1 in rat fibroblasts causes abnormalities in growth control, cell cycle progression and gene expression. *Oncogene* **8**, 3447–3457 (1993).

Acknowledgements

We thank R. Rothstein and L. Pons for helpful discussions, and J. Kitajewski and R. Dalla-Favera for providing reagents. This work was supported by the NIH, the American Cancer Society, an Irma T. Hirsch Trust Award and an Elaine B. Lesser Award in Ovarian Cancer Research.

Correspondence and requests for materials should be addressed to B.L. (e-mail: levine@cuccfa.ccc.columbia.edu).

Binding of paxillin to α_4 integrins modifies integrin-dependent biological responses

Shouchun Liu*, Sheila M. Thomas†, Darren G. Woodside*, David M. Rose*, William B. Kiosses*, Martin Pfaff†, & Mark H. Ginsberg*

*Department of Vascular Biology, VB-2, The Scripps Research Institute, 10550 N. Torrey Pines Rd, La Jolla, California 92037, USA

†Cancer Biology Program, Division of Hematology-Oncology, Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts 02215, USA

The α_4 integrins are indispensable for embryogenesis, haematopoiesis and immune responses^{1,2}, possibly because α_4 regulates cellular functions differently from other integrins through its cytoplasmic tail³. We used novel mimics⁴ of the α_4 tail to identify molecules that could account for α_4 -specific signalling. Here we report that the α_4 tail, but not several other α -subunit tails, binds tightly to the signalling adaptor paxillin. Paxillin physically associated with α_4 integrins in Jurkat T cells at high stoichiometry, and joining the α_4 tail to α_{1b} resulted in a complex of integrin $\alpha_{1b}\beta_3$ with paxillin. This association markedly enhanced the rates of $\alpha_{1b}\beta_3$ -dependent phosphorylation of focal adhesion kinase and cell migration. It also reduced cell spreading, focal adhesion and stress fibre formation. A point mutation within the α_4 tail that disrupts paxillin binding reversed all of these effects. Furthermore, $\alpha_4\beta_1$ -dependent adhesion to VCAM-1 led to spreading of mouse embryonic fibroblasts derived from paxillin-null but not from wild-type mice. Thus, the tight association of paxillin with the α_4 tail leads to distinct biochemical and biological responses to integrin-mediated cell adhesion.

To investigate the unusual signalling properties of α_4 integrins, we analysed the binding of cellular proteins to model protein mimics of dimerized α_4 integrin cytoplasmic domains. We made the model proteins by fusing the cytoplasmic tail of the α_4 or β_{1A} subunits to amino-terminal modules that contained four heptad-repeat sequences. The heptad repeats direct assembly of coiled-coil dimers in which the tails are dimerized parallel to each other and held in a fixed vertical stagger⁴, thus resembling the predicted topography of native integrin cytoplasmic domains⁵. Jurkat T-cell lysates were incubated with α_4 -tail mimics and bound proteins were detected by immunoblotting for integrin-associated polypeptides. Paxillin was enriched more than 57-fold in the bound fraction (Fig. 1a) relative to the cell lysate. In contrast, the β_{1A} tail bound paxillin with no enrichment relative to the cell lysate. There was no

‡ Present address: Ecole Normale Supérieure de Lyon, UMR 49, 46 allée d'Italie, 69364 Lyon, Cedex 07, France.

binding to empty resin or to the α_{IIb} tail. Heterodimers of $\alpha_4\beta_{1A}$ tails bound similar quantities of paxillin to the α_4 tail alone. The α_4 tail bound minimal quantities of filamin and talin (Fig. 1a) and no detectable vinculin or α -actinin (data not shown). In spite of seven conserved N-terminal residues in the α -subunits of integrin cytoplasmic tails^{5,6}, paxillin failed to bind to α_{IIb} , α_{3A} , α_5 or α_{6A} tails (Fig. 1b). Thus, paxillin binds specifically to the α_4 tail, and the shared^{5,6} N-terminal seven residues are not sufficient to mediate the interaction.

We examined the binding of paxillin-related proteins Hic-5 (ref. 7) and leupaxin⁸ to the α_4 tail. To assess Hic-5 binding, we used platelet⁹ extract; α_4 bound a 55K (relative molecular mass, 55,000) polypeptide which was reactive with an anti-paxillin that recognizes Hic-5 (ref. 9) (Fig. 1c). Jurkat cells contain leupaxin, which was also detected in the α_4 -bound and enriched fraction (Fig. 1d). Thus, paxillin and its two paralogues bind specifically to α_4 -tail affinity columns.

To exclude an indirect interaction, we examined the binding of purified recombinant paxillin. A paxillin-GST (glutathione S-transferase) fusion protein bound to the α_4 tail was saturable and of apparent high affinity (effector concentration at half-maximum response (EC_{50}), ~ 28 nM, Fig. 1e) using these recombinant proteins. This interaction did not require dimerized α_4 tail or GST-paxillin, as $\alpha_4\beta_1$ heterodimer, which is monomeric for α_4 , bound native paxillin (Fig. 1a). Thus, the α_4 tail binds paxillin directly and specifically.

The tight association of paxillin with the α_4 tail *in vitro* indicates that paxillin may bind α_4 integrins. To test this idea, we immunoprecipitated Jurkat cell lysate with antibodies against subunits of α_4 ,

β_1 or α_5 integrin. Paxillin was present in the α_4 and β_1 immunoprecipitates, but was not detected in those formed with anti- α_5 or irrelevant IgG (Fig. 2a). Analysis of surface proteins in the immunoprecipitates confirmed the activities of the three immunoprecipitating antibodies (Fig. 2a). We assessed the fraction of α_4 that was associated with paxillin by the sequential immunoprecipitation of surface-labelled Jurkat cell lysate with anti-paxillin. Paxillin depletion resulted in almost complete loss of α_4 in the lysate. In contrast, there was little depletion of α_5 (Fig. 2b). Immunoprecipitation with an irrelevant IgG did not result in significant loss of either $\alpha_4\beta_1$ or $\alpha_5\beta_1$ (Fig. 2b). In converse experiments, anti-paxillin immunoprecipitated $\alpha_4\beta_1$ but not $\alpha_5\beta_1$ (Fig. 2c). To exclude the possibility of a postlysis association, we mixed lysates formed from paxillin-null cells expressing α_4 with those formed from wild-type cells lacking α_4 . Immunoprecipitation from this mixed cell lysate failed to co-precipitate paxillin (data not shown). Thus the majority of native $\alpha_4\beta_1$ integrin can associate with endogenous cytoplasmic paxillin.

To determine whether the α_4 tail is sufficient to form a stable paxillin-integrin complex, we constructed a chimaera of the extra-cellular and transmembrane domains of α_{IIb} joined to the α_4 tail. For the appropriate β -tail partners, β_3 was joined to the β_{1A} or β_7 tail. The $\alpha_{IIb}\alpha_{6A}\beta_3\beta_{1A}$ and $\alpha_{IIb}\alpha_4\beta_3\beta_7$ chimaeric integrins were expressed in Chinese hamster ovary (CHO) cells; an $\alpha_{IIb}\alpha_{6A}\beta_3\beta_{1A}$ chimaera¹⁰ was used as a control. Antibodies against the extra-cellular domain of $\alpha_{IIb}\beta_3$ precipitated similar quantities of recombinant integrin from each cell line (Fig. 3a). Tenfold more paxillin was recovered with the α_4 -tail-bearing integrins. An α_4 -tail mutation, Y991A, which disrupts paxillin binding (S. Liu and M. H. Ginsberg, manuscript in preparation) was introduced into an $\alpha_{IIb}\alpha_4$

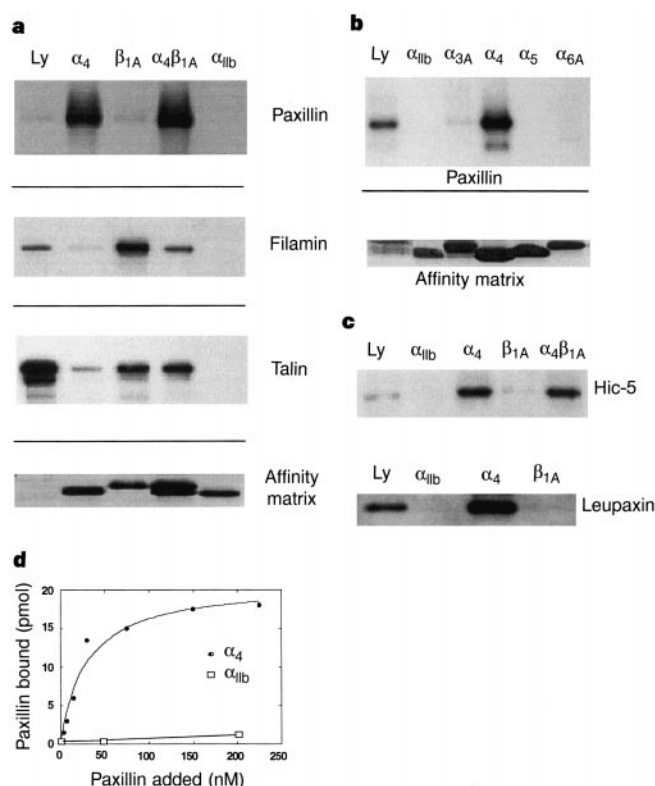


Figure 1 Preferential binding of paxillin to the α_4 tail (see also Supplementary Information). **a**, Identification by immunoblotting of the cellular polypeptides that bound to α_4 , β_{1A} , $\alpha_4\beta_{1A}$ or α_{IIb} tail mimics, using antibodies specific for paxillin, filamin or talin. Coomassie-blue staining of gels (affinity matrix) assessed the loading of each tail. **b**, Analysis of paxillin binding to different α integrin tails performed as in **a**. Ly, Jurkat cell lysate (**a**, **b**). **c**, Binding of Hic-5 (Ly, platelet cell lysate, upper panel) or leupaxin (Ly, Jurkat cell lysate, lower panel) to α_4 , β_{1A} , $\alpha_4\beta_{1A}$, or α_{IIb} tails. **d**, Identification of binding of purified recombinant paxillin to the α_4 tail.

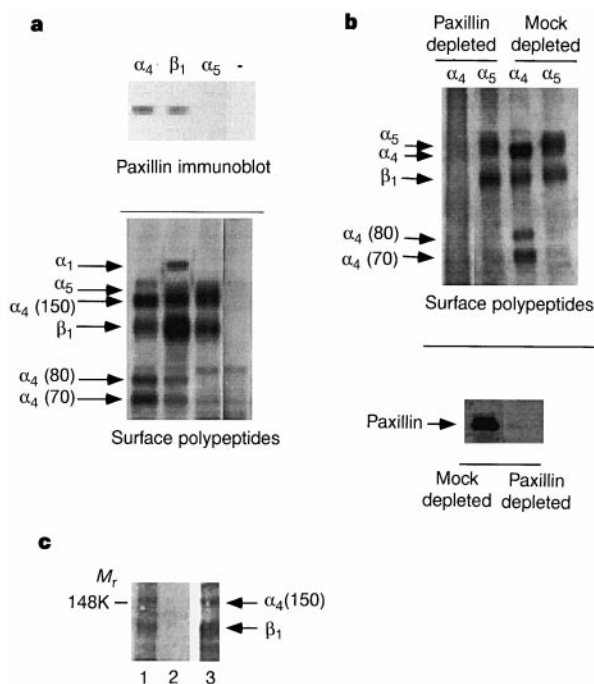


Figure 2 Association of paxillin with $\alpha_4\beta_1$ (see also Supplementary Information). **a**, Identification of paxillin co-precipitating with anti- α_4 (α_4), anti- β_1 (β_1), anti- α_5 (α_5), or an irrelevant mouse IgG (-) from a lysate of Jurkat T cells (upper panel). Immunoprecipitated cell-surface polypeptides using the same antibodies (lower panel). **b**, Cell-surface polypeptides immunoprecipitated by anti- α_4 or anti- α_5 (α_5) after multiple immunoprecipitations with anti-paxillin (paxillin depleted) or irrelevant IgG (mock depleted) from Jurkat cell lysate (upper panel). Absence of paxillin in paxillin-depleted or mock-depleted cell lysate (lower panel). **c**, Re-precipitation of $\alpha_4\beta_1$ from immunoprecipitates formed with anti-paxillin (1), irrelevant IgG (2) or anti- α_4 (3) from a Jurkat cell lysate.

chimaera. Roughly 12-fold less paxillin co-precipitated with the $\alpha_{IIB}\alpha_4(Y991A)\beta_3\beta_{1A}$ than with the wild-type $\alpha_{IIB}\alpha_4\beta_3\beta_{1A}$ chimaera (Fig. 3b). Thus, the α_4 tail mediates association of native integrins with paxillin.

The presence of the α_4 tail did not alter $\alpha_{IIB}\beta_3$ -dependent cell adhesion ($\alpha_{IIB}\alpha_4\beta_3\beta_{1A}$, $61 \pm 17\%$; $\alpha_{IIB}\alpha_6A\beta_3\beta_{1A}$, $66 \pm 16\%$ adhesion at 30 min) to fibrinogen, a ligand for integrin $\alpha_{IIB}\beta_3$. However, the α_4 tail opposed $\alpha_{IIB}\beta_3$ -dependent cell spreading (Fig. 4a) on fibrinogen. The $\alpha_4(Y991A)$ mutation restored $\alpha_{IIB}\beta_3$ -dependent cell spreading (Fig. 4a), but did not alter $\alpha_{IIB}\beta_3$ -dependent adhesion ($\alpha_{IIB}\alpha_4(Y991A)\beta_3\beta_{1A}$, $58 \pm 3\%$ adhesion). All cells spread to the same extent on fibronectin, a ligand for endogenous $\alpha_5\beta_1$ (data not shown). Furthermore, the $\alpha_4(Y991A)$ mutation also enhanced the spreading of Jurkat T cells on VCAM-1 (Fig. 4b). Thus, a point mutation that disrupts paxillin binding reduces the capacity of the α_4 tail to oppose cell spreading.

To confirm that paxillin is required for α_4 -mediated inhibition of cell spreading, we used a retroviral vector to express the α_4 subunit in fibroblasts derived from wild-type or paxillin-deficient mice, and assayed cell spreading on VCAM-1 (ref. 2). Primary mouse embryonic fibroblasts from two paxillin-null embryos spread on VCAM-1, whereas those from wild-type littermates failed to spread (Fig. 4c). We performed a similar experiment with immortalized α_4 -transfected fibroblasts to confirm that the effect was due to paxillin deficiency. Wild-type cells manifested reduced spreading compared with paxillin-null cells (data not shown). Furthermore, reconstruction of α_4 -expressing paxillin-null cells with paxillin resulted in loss of spreading when the cells were plated on VCAM-1 (Fig. 4c). Thus, paxillin is required for the diminished cell spreading induced by α_4 integrins.

Focal adhesion kinase (FAK) binds paxillin¹¹ and stimulates migration and focal adhesion (FA) disassembly^{12,13}. Because the α_4 tail enhances cell migration and reduces FAs¹⁴, we proposed that the binding of paxillin to α_4 integrins might change the kinetics of FAK tyrosine phosphorylation. When $\alpha_{IIB}\alpha_4\beta_3\beta_{1A}$ -bearing cells were attached to fibrinogen, a rapid and maximal FAK tyrosine phosphorylation occurred in 10 min. When the same cells were plated on fibronectin, a ligand for endogenous $\alpha_5\beta_1$, FAK phosphorylation was slower (maximal at 40–60 min) (Fig. 5a, b). In contrast, the rates of FAK phosphorylation on fibrinogen and fibronectin were similar when $\alpha_{IIB}\alpha_4(Y991A)\beta_3\beta_{1A}$ -bearing (Fig. 5a, b)

or $\alpha_{IIB}\alpha_6A\beta_3\beta_{1A}$ -bearing cells were used (data not shown). Furthermore, the α_4 tail did not cause a general increase in tyrosine phosphorylation because the rate of paxillin phosphorylation was similar during attachment of the $\alpha_{IIB}\alpha_4(Y991A)\beta_3\beta_{1A}$ -bearing and the $\alpha_{IIB}\alpha_4\beta_3\beta_{1A}$ -bearing cells to fibrinogen (data not shown). Thus, the α_4 tail changes the kinetics of FAK phosphorylation, and this change depends on the integrity of the paxillin-binding site in α_4 .

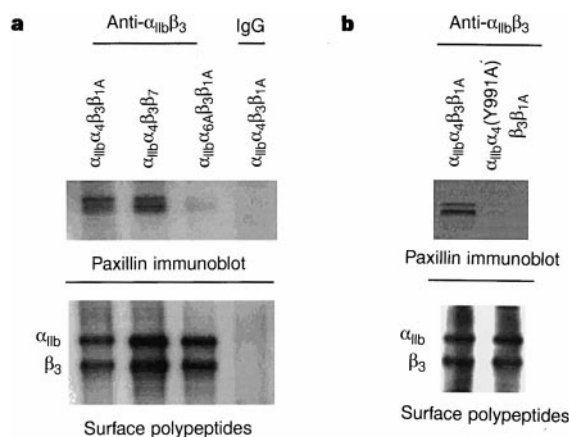


Figure 3 The α_4 tail increases the association of integrins with paxillin (see also Supplementary Information). **a**, Cell lysates from CHO cells stably expressing $\alpha_{IIB}\alpha_4\beta_3\beta_{1A}$, $\alpha_{IIB}\alpha_4\beta_3\beta_{1A}$, or $\alpha_{IIB}\alpha_6A\beta_3\beta_{1A}$ chimaeras were precipitated with $\alpha_{IIB}\beta_3$ -specific antibody, D-57 (anti- $\alpha_{IIB}\beta_3$) or an irrelevant IgG (IgG) (upper panel). Two paxillin bands were detected, probably representing the α and β isoforms of paxillin³⁰. Cell-surface proteins precipitated with anti- $\alpha_{IIB}\beta_3$ antibody or irrelevant IgG (lower panel). **b**, Paxillin co-precipitation with $\alpha_{IIB}\alpha_4\beta_3\beta_{1A}$ or $\alpha_{IIB}\alpha_4(Y991A)\beta_3\beta_{1A}$ chimaeric integrins (upper panel). Cell-surface polypeptides precipitated with anti- $\alpha_{IIB}\beta_3$ antibody (lower panel).

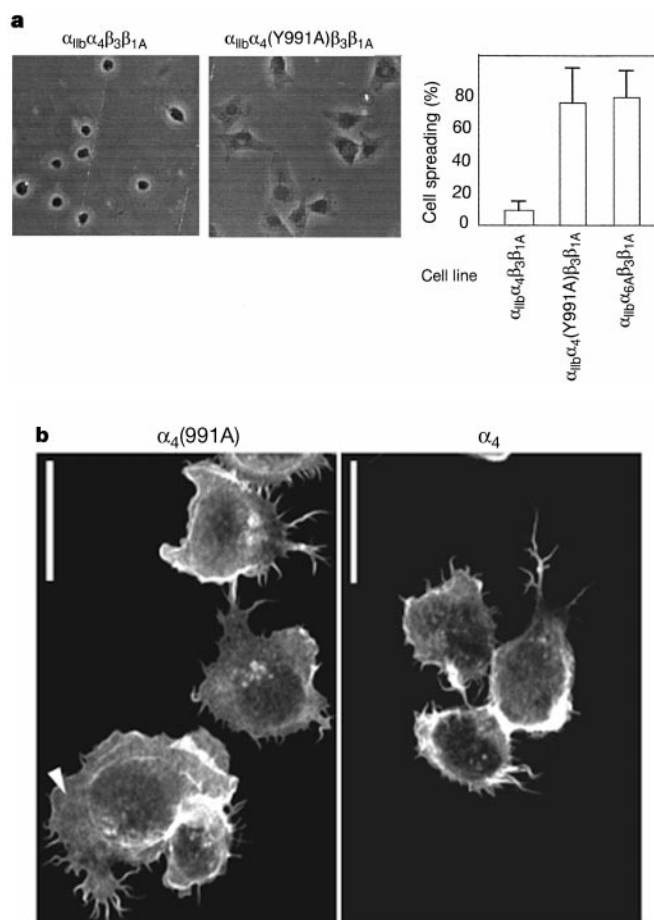


Figure 4 Association of the α_4 tail with paxillin inhibits cell spreading (see also Supplementary Information). **a**, Spreading on fibrinogen of CHO cells expressing $\alpha_{IIB}\alpha_4\beta_3\beta_{1A}$, $\alpha_{IIB}\alpha_4(Y991A)\beta_3\beta_{1A}$ or $\alpha_{IIB}\alpha_6A\beta_3\beta_{1A}$ chimaeras. Original magnification, $\times 200$. Data represent the mean $\pm$ s.d. **b**, Spreading of Jurkat T cells expressing wild-type α_4 or $\alpha_4(Y991A)$ integrin on VCAM-1. The mean cell area for $\alpha_4(Y991A)$ -expressing cells was 25.3% greater ($P < 0.001$) than that of wild-type α_4 -expressing cells on the basis of quantitative morphometry. **c**, Spreading of α_4 -expressing paxillin-null (paxillin-null, 1) or wild-type (wild-type, 2) mouse embryonic fibroblasts on VCAM-1. Transfection of paxillin-null cells with a complementary DNA encoding human paxillin reversed α_4 -dependent cell spreading. Data represent the mean $\pm$ s.d.

We used the $\alpha_{IIb}\beta_3$ chimaeras to determine whether the paxillin-binding site is also required for the effects of the α_4 tail on FA formation and cell migration¹⁴. $\alpha_{IIb}\alpha_4\beta_3\beta_{1A}$ -bearing cells remained predominantly rounded and manifested few stress fibres (Fig. 6a, b) during attachment to fibrinogen. About 20% of the cells were polarized and exhibited lamellipodia and talin-containing complexes; however, these complexes were not usually associated with stress fibres (Fig. 6b). In contrast, the $\alpha_{IIb}\alpha_4(Y991A)\beta_3\beta_{1A}$ -bearing cells (Fig. 6b) were spread and contained multiple stress fibres terminating in typical FAs (Fig. 6a, b). On the fibronectin, both cell lines manifested similar stress fibre and FA formation (Fig. 6a). Furthermore, $\alpha_{IIb}\alpha_4\beta_3\beta_{1A}$ -bearing cells migrated more rapidly on fibrinogen than those bearing $\alpha_{IIb}\alpha_4(Y991A)\beta_3\beta_{1A}$ (Fig. 6c). Consequently, the paxillin-binding site on the α_4 tail is required to promote cell migration and oppose FA formation.

The paxillin- α_4 -tail association is strong compared with previously described¹⁵ interactions of integrin tails. The strength of this association is evidenced by the marked enrichment of paxillin by α_4 -tail affinity chromatography and stoichiometric co-precipitation of α_4 integrins with paxillin in a relatively hydrophobic detergent (1% Triton X-100). Furthermore, the paxillin- α_4 association appears specific for α_4 among integrin α tails and contributes to the unusual biological effects of the α_4 tail¹⁴. An α_4 -tail mutation that blocks paxillin binding or lack of paxillin expression promotes α_4 -integrin-dependent cell spreading. In addition, loss of paxillin binding to the α_4 tail results in increased FA formation and reduced cell migration. Thus, binding of paxillin to the α_4 tail leads to altered cytoskeletal organization and cell migration.

The α_4 tail modulates the kinetics of integrin signalling through its paxillin-binding site. Cell spreading, FA formation and cell migration require coordinated signalling between multiple biochemical

pathways¹⁶. An important component of these pathways is FAK, a paxillin-binding tyrosine kinase¹¹. As the paxillin-binding site of the α_4 tail was required for the α_4 -tail-induced increase in the initial rate of FAK phosphorylation, and as paxillin and FAK are physically associated and membrane localization and clustering of FAK are involved in its phosphorylation and activation^{17,18}, we propose that the α_4 -paxillin interaction promotes FAK phosphorylation by

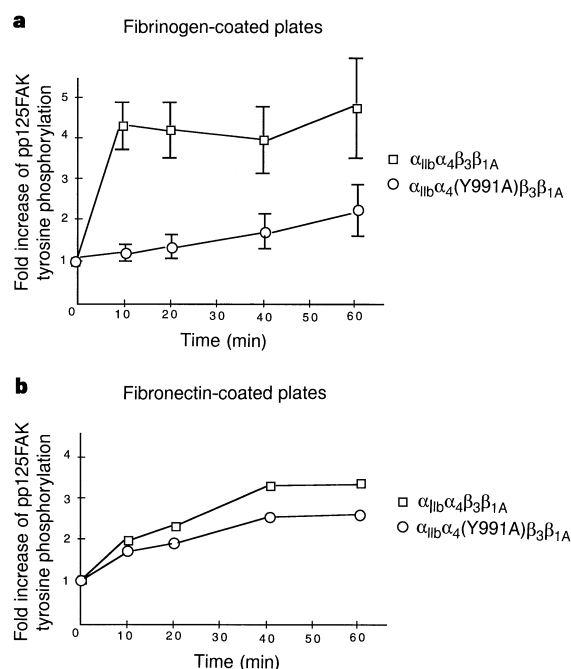


Figure 5 The α_4 tail changes the kinetics of FAK phosphorylation (see also Supplementary Information). **a**, $\alpha_{IIb}\alpha_4\beta_3\beta_{1A}$ - or $\alpha_{IIb}\alpha_4(Y991A)\beta_3\beta_{1A}$ -expressing CHO cells were plated on fibrinogen-coated plates for the indicated durations. Anti-FAK antibody immunoprecipitates were fractionated by SDS-PAGE, and FAK phosphorylation was detected using PY20 and 4G10 antibodies. Equal loading of FAK protein was verified by immunoblotting of the stripped blots with anti-FAK antibody (data not shown). Data are means $\pm$ s.e.m. of three independent experiments. **b**, CHO cells expressing $\alpha_{IIb}\alpha_4\beta_3\beta_{1A}$ or $\alpha_{IIb}\alpha_4(Y991A)\beta_3\beta_{1A}$ were plated on fibronectin-coated plates. Immunoprecipitation and FAK phosphorylation were performed as described in **a**. Data are means of duplicate experiments.

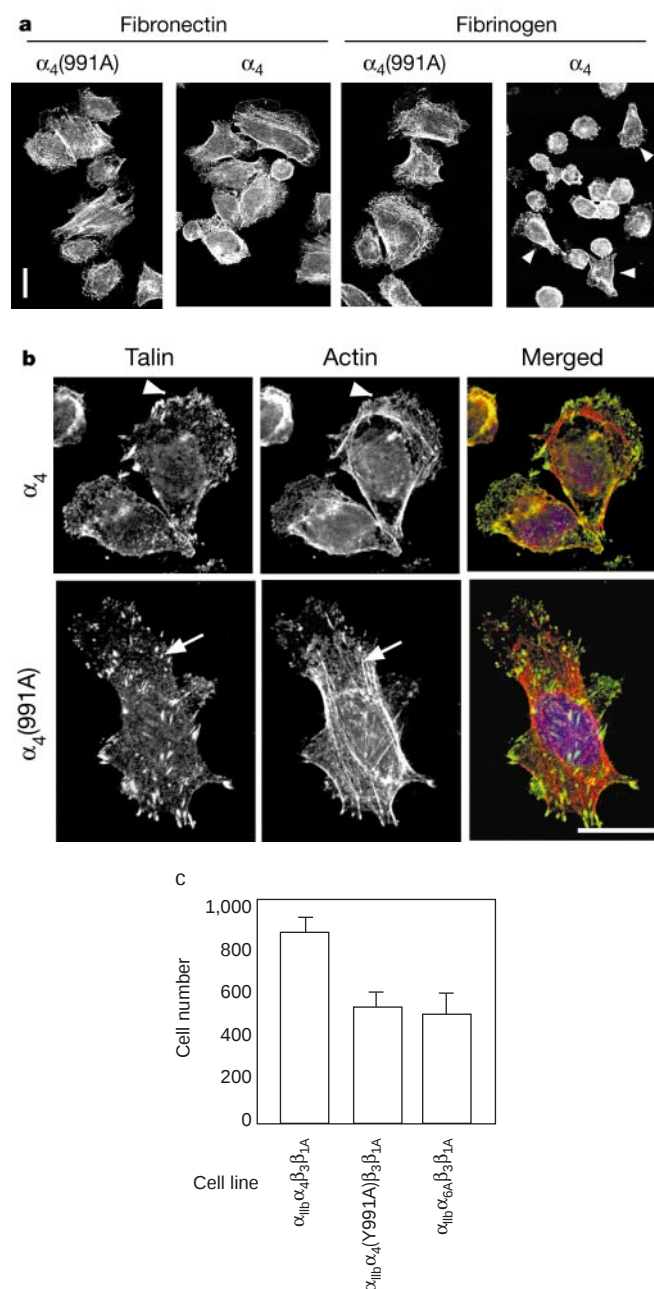


Figure 6 Inhibition of FA formation and enhancement of cell migration by the α_4 tail depends on its paxillin-binding site (see also Supplementary Information). **a**, **b**, CHO cells expressing $\alpha_{IIb}\alpha_4\beta_3\beta_{1A}$ or $\alpha_{IIb}\alpha_4(Y991A)\beta_3\beta_{1A}$ bound to fibrinogen or fibronectin were stained with rhodamine-phalloidin to detect F-actin, and were also stained with anti-talin in **b**. CHO cells expressing $\alpha_{IIb}\alpha_4\beta_3\beta_{1A}$ displayed diminished spreading, stress fibres and FAs on fibrinogen. About 20% of the $\alpha_{IIb}\alpha_4\beta_3\beta_{1A}$ cells showed a polarized morphology (arrowheads in **a**). These cells manifested lamellipodia with focal accumulation of talin and few stress fibres (arrowheads in **b**). CHO cells expressing $\alpha_{IIb}\alpha_4(Y991A)\beta_3\beta_{1A}$ spread and formed stress fibres and FAs (arrows in **a**, **b**) on fibrinogen. Both types of cell spread and formed stress fibres on fibronectin, a ligand for endogenous integrin $\alpha_5\beta_1$ (**a**). **c**, Random migration on fibrinogen of CHO cells expressing $\alpha_{IIb}\alpha_4\beta_3\beta_{1A}$, $\alpha_{IIb}\alpha_4(Y991A)\beta_3\beta_{1A}$, or $\alpha_{IIb}\alpha_6\beta_3\beta_{1A}$. Data are means $\pm$ s.d. of triplicate experiments.

increasing the membrane targeting and clustering of FAK. FAK phosphorylation is involved in its activation and in its coupling to Src kinases^{17–19}. Furthermore, FAK and Src family kinases are essential for integrin-dependent cell migration and for disassembly of FAs²⁰. Consequently, alteration in the kinetics of integrin signalling may account for the effects of the paxillin- α_4 -tail interaction on cell migration and FA formation.

The paxillin- α_4 -tail association may influence many signalling events in addition to FAK phosphorylation. Paxillin binds, and can be a substrate for, protein tyrosine kinases, such as Pyk2, Src and Abl (ref. 11). Paxillin also binds p95PKL (ref. 21), which can interact with a complex containing the guanine nucleotide exchange factor, PIX, the adapter, Nck, and the serine/threonine kinase, PAK. Furthermore, during integrin-mediated cell adhesion or in response to growth factor stimulation, paxillin can be tyrosine and serine phosphorylated¹¹. Tyrosine phosphorylation of paxillin forms docking sites for the SH-2 domains of proteins such as Crk, with consequent recruitment of guanine nucleotide exchangers such as C3G and Sos¹¹. Many paxillin binding proteins, including FAK¹², Src-family kinases¹⁹, PTP-PEST^{13,22,23} and Crk²⁴, have been implicated in stimulating cell migration and FA disassembly. These paxillin-associated proteins can also regulate cell survival and proliferation. Thus, our results provide a new insight into α_4 -specific signal transduction and suggest that the paxillin- α_4 -tail interaction may regulate multiple cellular functions by modifying the kinetics of integrin signalling. □

Methods

Cell lines and culture

Human Jurkat T cells and CHO cell lines were cultured as described⁴. Human $\alpha_{IIb}\beta_3$ (or $\alpha_{IIb}\beta_3$ (Y991A)) and $\beta_3\beta_{1A}$ (or $\beta_3\beta_7$) chimaeras were formed by connecting α_{IIb} and β_3 extracellular and transmembrane domains to α_4 (or α_4 (Y991A)) and β_{1A} (or β_7) tails, respectively. CHO cells stably expressing these chimaeras were transfected and isolated as described¹⁰. Primary mouse embryonic fibroblasts from paxillin-null and littermate-matched wild-type embryos were isolated as described (M. Hagel *et al.* manuscript in preparation²⁵).

Integrin tail model proteins

The design, production and purification of tail model protein has been described⁴. $\alpha_4\beta_{1A}$ heterodimer model protein was prepared by oxidation of equimolar mixture of α_4 and β_{1A} model proteins, and isolated by reverse phase C₁₈ HPLC.

Affinity chromatography

Cell lysate from Jurkat T cells or platelets was prepared and used for tail-affinity chromatography as described⁴. The bound proteins were detected by western blotting. Expression and isolation of recombinant human paxillin have been described²⁶. Aliquots of GST-paxillin or GST in 300 μ l binding buffer⁴ plus 1 mg ml⁻¹ of BSA were added to resins loaded with model proteins, and incubated at room temperature for 2 h. Bound proteins were detected with antibodies for HA-tag (12CA5, ATCC) or GST (B-14, Santa Cruz).

Immunoprecipitation and western blotting

Jurkat T cells or CHO cells were surface labelled with sulfo-NHS-Biotin (Pierce). Cell lysate was prepared and immunoprecipitation performed as described²⁷. Precipitates were separated under non-reducing conditions and detected with streptavidin-peroxidase and ECL. Co-precipitation of paxillin was detected similarly except that cells were not surface biotinylated and immunoprecipitates were separated under reducing conditions and paxillin co-precipitation was detected with biotin-labelled anti-paxillin (prepared by labelling anti-paxillin (clone 349, Transduction Laboratories) with NHS-Biotin). For paxillin-depletion assay, aliquots of surface biotinylated Jurkat cell lysate were subjected to immunoprecipitation using anti-paxillin or irrelevant IgG. Western blotting assessed the degree of paxillin-depletion in cell lysates. Cell lysates with or without paxillin-depletion, and with irrelevant IgG precipitation, were immunoprecipitated with either anti- α_4 (HP2/1, Immuno Tech) or anti- α_5 (PharMingen) antibody. Immunoprecipitates were separated under non-reducing conditions by SDS-PAGE and surface polypeptides were detected with streptavidin-peroxidase. For reciprocal anti-paxillin immunoprecipitation, surface biotinylated Jurkat cell lysate was immunoprecipitated with anti-paxillin or irrelevant IgG. The precipitates were extracted with non-reducing SDS sample buffer, diluted with IP buffer, precipitated with anti- α_4 antibody and analysed by SDS-PAGE.

Cell adhesion and spreading

Cell adhesion and spreading were performed as described²⁸. All Paxillin-null and wild-type mouse cells were infected with a pBABE retroviral vector encoding human α_4 integrin and equal expression of α_4 integrins was routinely verified by FACS analysis (data not shown). Cells resuspended in DMEM plus 1% BSA were plated on coverslips coated with

10 μ g ml⁻¹ either VCAM-1 or fibronectin and spreading at 37 °C was assessed by phase microscopy after 1 h. Immortalized paxillin-null and wild-type cells were obtained by transformation with SV40 large T antigen and were transfected with α_4 . For paxillin reconstitution, α_4 -expressing paxillin-null cells were transiently transfected with vectors encoding paxillin or empty vector together with vector encoding green fluorescent protein (GFP). Spreading of GFP(+) cells was assessed by fluorescence microscopy after 48 h.

Tyrosine phosphorylation

Petri dishes (100 mm) were coated with 10 μ g ml⁻¹ fibrinogen or fibronectin and blocked with heat-denatured BSA. Cells were resuspended in DMEM plus 1% BSA at 37 °C with rotation for 3–4 h, and then added to coated petri dishes and incubated at 37 °C. At each time interval, a set of dishes was washed, and adhered cells were extracted with radio-immunoprecipitation assay (RIPA) buffer. Cell lysate was immunoprecipitated with anti-FAK²⁹ or anti-paxillin. Tyrosine phosphorylation was detected by western blotting using anti-phosphotyrosine antibodies PY-20 and 4G10 (Transduction Laboratories). The membrane was stripped, and re-blotted with anti-FAK or anti-paxillin to verify similar recoveries.

Immunofluorescence and cell migration assays

Glass coverslips were coated with 10 μ g ml⁻¹ fibrinogen, fibronectin or VCAM-1. Cells were plated on coated coverslips for 45 min, fixed, permeabilized, and stained with anti-talin monoclonal antibody (clone 8d4, Sigma) and FITC-conjugated goat-anti-mouse (BioSource), rhodamine-phalloidin and TOTO-3 (Molecular Probes). Cell migration was performed using Transwell chambers (8 μ m, Costar). Both sides of chambers were coated with 10 μ g ml⁻¹ fibrinogen overnight at 4 °C. The coated chambers were blocked with heat-denatured BSA. Cells resuspended in DMEM plus 1% BSA were added to the upper chamber and incubated at 37 °C for 3 h. The cells that migrated to the lower side were stained with 0.5% Crystal Violet (Sigma) and enumerated by microscopy.

Received 6 August; accepted 19 October 1999.

- Stewart, M., Thiel, M. & Hogg, N. Leukocyte integrins. *Curr. Opin. Cell Biol.* **7**, 690–696 (1995).
- Shimizu, Y., Rose, D. M. & Ginsberg, M. H. Integrins in the immune system. *Adv. Immunol.* **72**, 325–380 (1999).
- Hemler, M. E., Kassner, P. D. & Chan, B. M. C. Functional roles for integrin alpha subunit tails. *Cold Spring Harbor Symposia on Quant. Biol.* **57**, 213–220 (1992).
- Pfaff, M., Liu, S., Erle, D. J. & Ginsberg, M. H. Integrin β tails differentially bind to cytoskeletal proteins. *J. Biol. Chem.* **273**, 6104–6109 (1998).
- Williams, M. J., Hughes, P. E., O'Toole, T. E. & Ginsberg, M. H. The inner world of cell adhesion: integrin tails. *Trends Cell Biol.* **4**, 109–112 (1994).
- Sastry, S. K. & Horwitz, A. F. Integrin tails: mediators of cytoskeletal linkages and extra- and intracellular initiated transmembrane signaling. *Curr. Opin. Cell Biol.* **5**, 819–831 (1993).
- Shibanuma, M. *et al.* Induction of senescence-like phenotypes by forced expression of the hic-5, which encodes a novel LIM motif protein, in immortalized human fibroblasts. *Mol. Cell. Biol.* **17**, 1224–1235 (1994).
- Lipsky, B. P., Beals, C. R. & Staunton, D. E. Leupaxin is a novel LIM domain protein that forms a complex with PYK2. *J. Biol. Chem.* **273**, 11709–11713 (1998).
- Hagmann, J., Grob, M., Welman, A., van Willigen, G. & Burger, M. M. Recruitment of the LIM protein hic-5 to focal contacts of human platelets. *J. Cell Sci.* **11**, 2181–2188 (1998).
- Hughes, P. E. *et al.* Suppression of integrin activation: a novel function of a Ras/Raf-initiated MAP kinase pathway. *Cell* **88**, 521–530 (1997).
- Turner, C. E. Paxillin. *Int. J. Biochem. Cell Biol.* **30**, 955–959 (1998).
- Ilic, D. *et al.* Reduced cell motility and enhanced focal adhesion contact formation in cells from FAK-deficient mice. *Nature* **377**, 539–544 (1995).
- Cary, L. A., Chang, J. F. & Guan, J.-L. Stimulation of cell migration by overexpression of focal adhesion kinase and its association with Src and Fyn. *J. Cell Sci.* **109**, 1787–1794 (1996).
- Kassner, P. D., Alon, R., Springer, T. A. & Hemler, M. Specialized functional properties of the integrin α_4 tail. *Mol. Biol. Cell* **6**, 661–674 (1995).
- Hemler, M. E. Integrin associated proteins. *Curr. Opin. Cell Biol.* **10**, 578–585 (1998).
- Lauffenburger, D. A. & Horwitz, A. F. Cell migration: a physically integrated molecular process. *Cell* **84**, 359–369 (1996).
- Parsons, J. T. & Parsons, S. J. Src family protein tyrosine kinases: cooperating with growth factor and adhesion signaling pathways. *Curr. Opin. Cell Biol.* **9**, 187–192 (1997).
- Hanks, S. K. & Polte, T. R. Signaling through focal adhesion kinase. *Bioessays* **19**, 137–145 (1997).
- Klinghoffer, R. A., Sachsenmaier, C., Cooper, J. A. & Soriano, P. Src family kinases are required for integrin but not PDGFR signal transduction. *EMBO J.* **18**, 2459–2471 (1999).
- Fincham, V. J. & Frame, M. C. The catalytic activity of Src is dispensable for translocation to focal adhesions but controls the turnover of these structures during cell motility. *EMBO J.* **17**, 81–92 (1998).
- Turner, C. E. *et al.* Paxillin LD4 motif binds PAK and PIX through a novel 95-kD ankyrin repeat, ARF-GAP protein: a role in cytoskeletal remodeling. *J. Cell Biol.* **145**, 851–863 (1999).
- Shen, Y., Schneider, G., Cloutier, J. F., Veillette, A. & Schaller, M. D. Direct association of protein-tyrosine phosphatase PTP-PEST with paxillin. *J. Biol. Chem.* **273**, 6474–6481 (1998).
- Angers-Loustau, A. *et al.* Protein tyrosine phosphatase-PEST regulates focal adhesion disassembly, migration, and cytokinesis in fibroblasts. *J. Cell Biol.* **144**, 1019–1031 (1999).
- Klemke, R. L. *et al.* CAS/Grb2 coupling serves as a “molecular switch” for induction of cell migration. *J. Cell Biol.* **140**, 961–972 (1998).
- Thomas, S. M., Soriano, P. & Imamoto, A. Specific and redundant roles of Src and Fyn in organizing the cytoskeleton. *Nature* **376**, 267–271 (1995).
- Salgia, R. *et al.* Molecular cloning of human paxillin, a focal adhesion protein phosphorylated by P210BCR/ABL. *J. Biol. Chem.* **270**, 5039–5047 (1995).
- Chen, Y.-P., O'Toole, T. E., Ylanne, J., Rosa, J.-P. & Ginsberg, M. H. A point mutation in the integrin β_3 tail (S752-P) impairs bidirectional signaling through $\alpha_{IIb}\beta_3$ (Platelet glycoprotein IIb-IIIa). *Blood* **84**, 1857–1865 (1994).
- Ylanne, J. *et al.* Distinct functions of integrin α and β subunit tails in cell spreading and formation of focal adhesions. *J. Cell Biol.* **122**, 223–233 (1993).

29. Gao, J., Zoller, K. E., Ginsberg, M. H., Brugge, J. S. & Shattil, S. J. Regulation of the pp72syk protein tyrosine kinase by platelet integrin α IIb β 3. *EMBO J.* **16**, 6414–6425 (1997).
 30. Mazaki, Y., Uchida, H., Hino, O., Hashimoto, S. & Sabe, H. Paxillin isoforms in mouse. *J. Biol. Chem.* **273**, 22435–22441 (1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

The authors thank R. Salgia and J. Griffin for the gift of vectors encoding paxillin, and S. Shattil for critical review of the manuscript. This work was supported by grants from the NIH.

Correspondence and requests for materials should be addressed to M.H.G. (e-mail: ginsberg@scripps.edu).

A novel nuclear export activity in HIV-1 matrix protein required for viral replication

Stefan Dupont^{††}, Natalia Sharova[†], Caryn DéHoratius[†], Ching-Man A. Virbasius[†], Xiaochun Zhu[†], Alice G. Bukrinskaya[†], Mario Stevenson[†] & Michael R. Green^{††}

^{*}Howard Hughes Medical Institute and ^{††}Program in Molecular Medicine at the University of Massachusetts Medical Center, 373 Plantation Street, Suite 309, Worcester, Massachusetts 01605, USA

An important aspect of the pathophysiology of human immunodeficiency virus type-1 (HIV-1) infection is the ability of the virus to replicate in non-dividing cells^{1–3}. HIV-1 matrix (MA), the amino-terminal domain of the Pr55 gag polyprotein (Pr55), bears a nuclear localization signal that promotes localization of the viral preintegration complex to the nucleus of non-dividing cells following virus entry^{3–5}. However, late during infection, MA, as part of Pr55, directs unspliced viral RNA to the plasma membrane⁶, the site of virus assembly. How MA can mediate these two opposing targeting functions is not understood. Here we demonstrate that MA has a previously undescribed nuclear export activity. Although MA lacks the canonical leucine-rich nuclear export signal, nuclear export is mediated through the conserved Crmlp pathway and functions in both mammalian cells and yeast. A mutation that disrupts the MA nuclear export signal (MA-M4) mislocalizes Pr55 and genomic viral RNA to the nucleus, thereby severely impairing viral replication. Furthermore, we show that MA-M4 can act in a dominant-negative fashion to mislocalize genomic viral RNA even in the presence of wild-type MA. We conclude that the MA nuclear export signal is required to counteract the MA nuclear localization signal, thus ensuring the cytoplasmic availability of the components required for virion assembly.

We have carried out a systematic mutational analysis of MA within the X4 tropic infectious molecular clone, HIV-1_{HXB2} (ref. 7). One such mutant, HXB2-M4, which comprises substitution of two basic residues at positions 18 and 22 of MA (Fig. 1a), is the focus of this study. The spreading infection assay of Fig. 1b and single cycle infectivity assay of Fig. 1c show that HXB2-M4 is severely replication-defective. To determine the basis of this replication defect we analysed known properties and functions of MA (see ref. 8). We found that Pr55-M4 underwent normal proteolytic processing (Fig. 1d), was properly myristoylated (Fig. 1e), and supported viral envelope incorporation (Fig. 1f). Figure 1g shows that HXB2-M4 virions had a reduced ratio of genomic viral RNA to gag p24 (lanes 4 and 5) compared with the wild-type counterpart

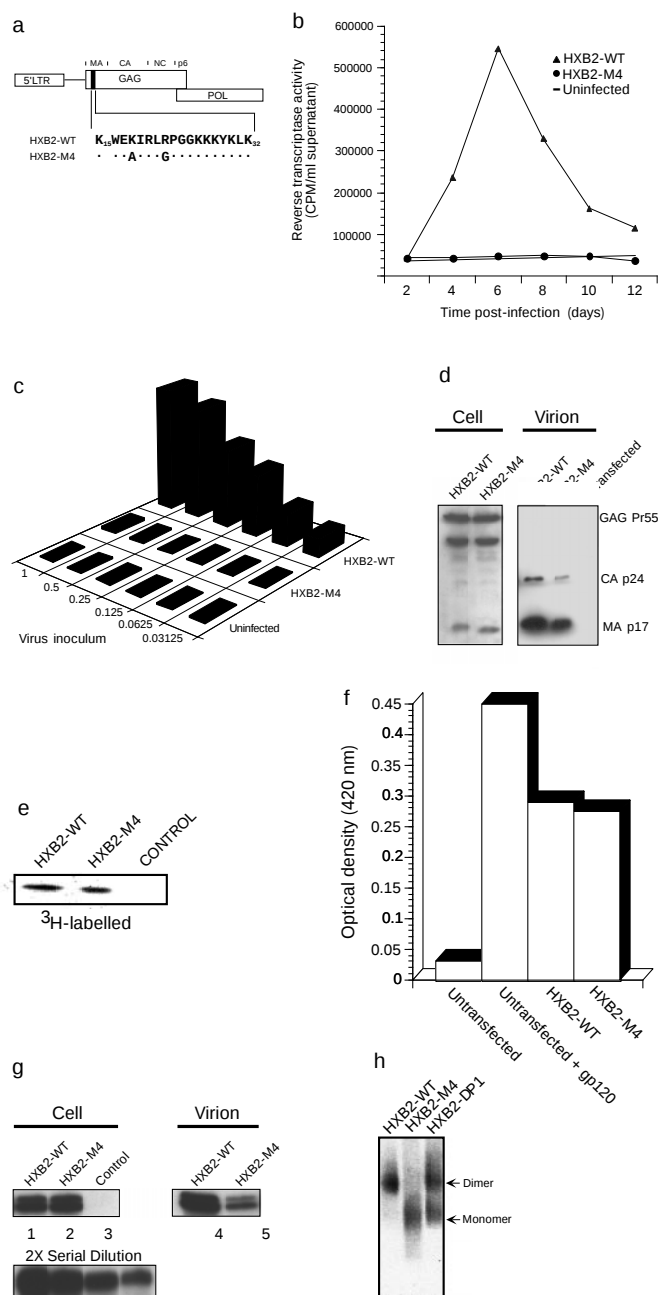


Figure 1 Characterization of a novel HIV-1 MA mutant. **a**, Schematic diagram of HXB2-M4. **b**, HXB2-M4 is replication-defective. Wild-type and mutant virus replication in MT-4 cells was monitored by reverse transcriptase activity in culture supernatants at 2-day intervals ($n = 3$). **c**, Single cycle infectivity assay. HeLa-CD4-LTR/ β -gal Magi indicator cells (2×10^4 cells per well) were infected with HIV-1 virions (normalized to equivalent gag p24 levels) and β -galactosidase production was determined at 48 h post-infection. **d**, Proteolytic processing of Pr55-M4. Immunoblots of total cellular lysates from virus producing cells (left) and purified virions (right) with a mixture of antibodies to gag MA and gag CA. **e**, Myristoylation of HIV-1 MA. Viral producer cells were supplemented with ³H-Myristic acid and gag MA was immunoprecipitated with an anti-gag MA monoclonal antibody and visualized by autoradiography. **f**, Virion incorporation of gp120. Gradient-purified virions were normalized to equivalent gag p24 levels and virion associated envelope glycoprotein was measured with a gp120 ELISA kit. **g**, RNase protection assay. 10 µg of total cellular RNA (left) and virion-associated RNA (right) obtained from virus particles (350 ng p24 equivalents) were analysed. RNA from untransfected cells was used as a control. **h**, Dimerization assay. Equivalent amounts of wild-type and mutant virions, based on genomic RNA content, were examined. HXB2-ΔP1 contains a 19-bp deletion within the dimerization initiation sites and packages monomers and dimers^{9,10}.

despite equivalent levels of genomic RNA within cells (lanes 1 and 2). We noted that the small amount of genomic RNA incorporated into HXB2-M4 virions was in the form of a monomer rather than a dimer (Fig. 1h); the dimeric form is required for proper reverse transcription and infectivity^{9,10}. We conclude that the replication-defect of HXB2-M4 is related to an inability to correctly package genomic viral RNA into virions.

To understand the basis for the decreased viral RNA content of virions, we determined the cellular localization of genomic viral RNA by *in situ* hybridization. Figure 2A shows that in cells producing wild-type virus (HXB2-WT), viral RNA was predominantly cytoplasmic (panel a), whereas in cells that were producing HXB2-M4 viral RNA was predominantly nuclear (panel b). The mislocalization defect resembled (but was not as severe as) the defect

occurring in a Rev-defective mutant (panel c). These results indicate that the viral RNA packaging defect results, at least in part, from the nuclear mislocalization of genomic viral RNA.

Figure 2B shows that the mislocalization defect was specific to viral RNA: total poly(A)⁺ RNA was localized normally in cells producing HXB2-M4. Interestingly, Fig. 2C shows that the viral RNA mislocalization defect was evident at late (24 hours) but not at early (12 hours) times post-transfection. By comparison, viral RNA was mislocalized in a Rev-defective mutant at both early and late times post-transfection (panels c, f).

We tested whether Pr55 might be involved in nuclear accumulation of viral RNA by analysing the cellular localization of wild-type and mutant Pr55 derivatives after fusion to green fluorescent protein (GFP) (Fig. 3A). Figure 3B shows, as expected, that GFP itself was diffusely localized throughout both the nucleus and cytoplasm, Pr55-GFP was cytoplasmic, and Pr55/M4-GFP was predominantly nuclear. Collectively, these results indicate, first, that the M4 mutation disrupts the MA cytoplasmic localization signal and, second, consistent with previous studies³⁻⁵, Pr55 must also contain a nuclear localization signal (NLS). A Pr55 variant containing a MA deletion (Pr55ΔMA-GFP) was cytoplasmic, confirming that the MA portion of Pr55 contained the NLS.

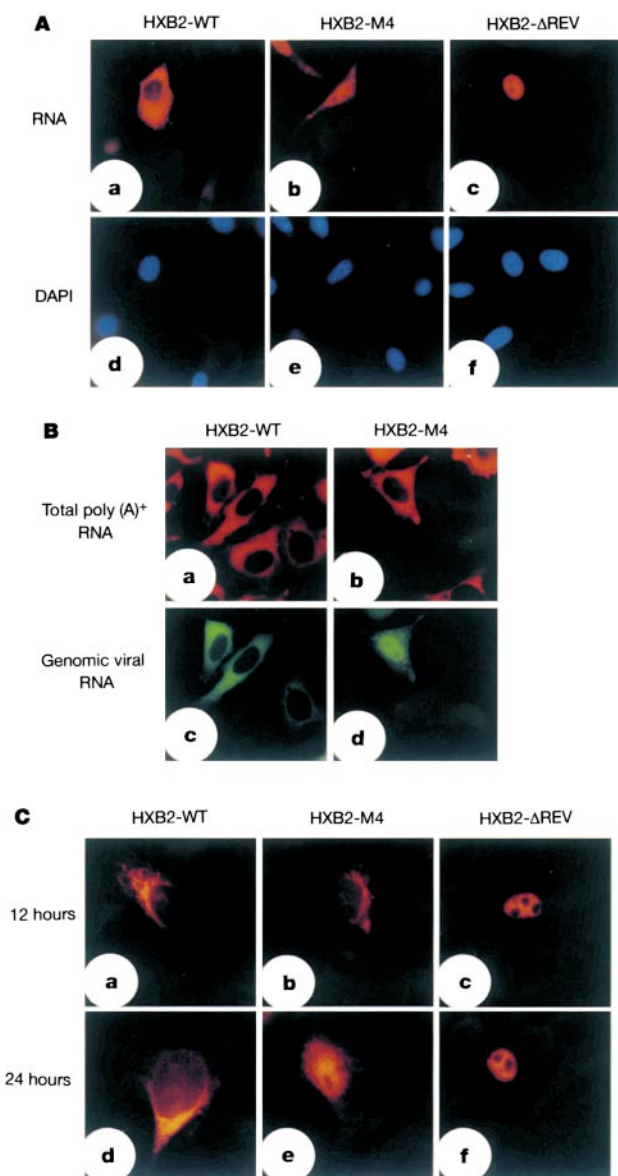


Figure 2 Cellular distribution of viral and total poly(A)⁺ RNA by *in situ* hybridization. **A**, Genomic viral RNA. Unspliced viral RNA was detected with a Cy3-labelled antisense probe corresponding to a 91-bp segment (HIV-1_{HXB2} 2553–2462) (panels a–c). Nuclei were stained with 6-diamidino-2-phenylindole (DAPI) (panels d–f). **B**, Total poly(A)⁺ RNA. Cellular distribution of poly(A)⁺ RNA (panels a, b) and unspliced HIV-1 RNA (panels c, d). Poly(A)⁺ RNA was detected with an oligonucleotide (dT)₅₀ digoxigenin-labelled probe. **C**, Time-course. Distribution of viral genomic RNA was examined at 12 h (panels a–c) and 24 h (panels d–f) post-transfection with the indicated proviral clones.

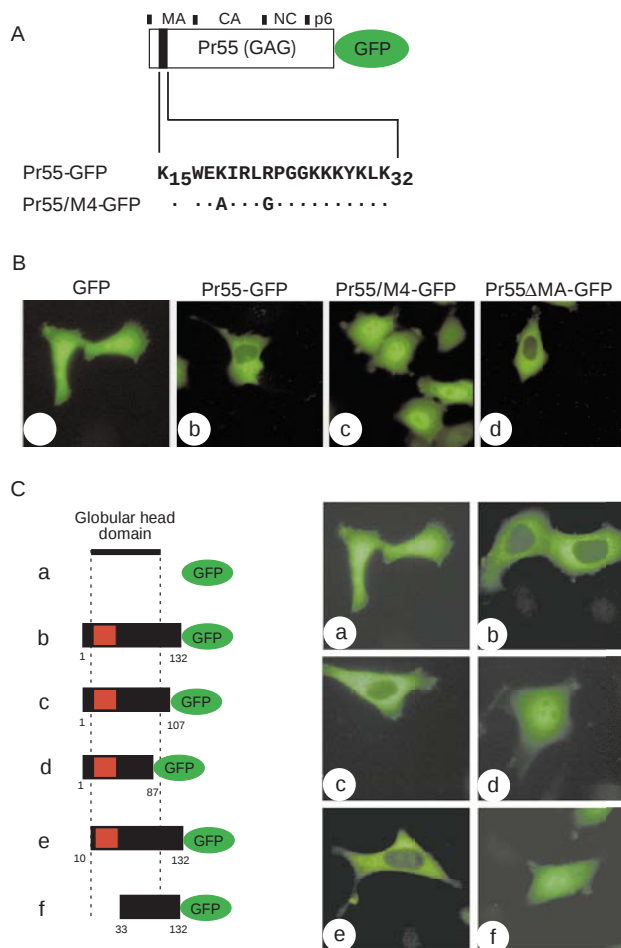


Figure 3 Cellular distribution of Pr55 and MA variants. **A**, Schematic diagram of Pr55 and Pr55/M4-GFP fusion constructs. **B**, Cellular localization of Pr55-GFP and Pr55/M4-GFP. Cellular localization of GFP (panel a), Pr55-GFP (panel b), Pr55/M4-GFP (panel c) and Pr55ΔMA-GFP (panel d) was determined by immunofluorescence microscopy. **C**, Cellular localization of truncated MA variants. Schematic diagrams of MA derivatives are shown on the left. Letters a–f correspond to panels a–f on the right. A series of positively charged residues (15–33; red box) reside near the N terminus of MA and encompass the previously identified NLS (ref. 4).

To map the novel cytoplasmic localization signal, we analysed a series of N- and C-terminal MA deletion mutants. Both N- and C-terminal deletions that encroach on the highly structured 'globular head domain'^{11,12} disrupted the cytoplasmic localization signal (Fig. 3C).

In principle, a cytoplasmic localization signal could function either by cytoplasmic retention or nuclear export. Leptomycin B is a potent inhibitor of mammalian Crm1p (refs 13,14) a central component in the protein nuclear export pathway (refs 15–17). Figure 4A shows that MA-GFP became nuclear-localized following inhibition of nuclear export with Leptomycin B (panel c).

The Crm1p pathway is highly conserved from yeast to man raising the possibility that the MA nuclear export activity would also function in yeast. To test this possibility we used a previously described yeast transport assay involving a large GFP fusion-protein, in order to prevent nuclear–cytoplasmic diffusion, and containing the SV40 Large-T NLS, in order to direct nuclear localization. To this reporter we added either the wild-type (MA-WT) or mutant (MA-M4) protein (Fig. 4B). Figure 4C shows that wild-type MA remained cytoplasmic (panel a), whereas MA-M4 accumulated within the nucleus (panel c). Thus, the MA nuclear export signal (NES) works in human cells as well as in yeast cells.

The yeast strain *xpo1-1* contains a temperature-sensitive mutation within *XPO1* (ref. 18), the yeast homologue of the mammalian *CRM1*. Figure 4D shows that after temperature-sensitive inactivation of Xpo1p, MA became nuclear-localized (panel c). We conclude that the HIV-1 MA NES functions in yeast cells, as well as in mammalian cells, and in both cases is dependent upon Crm1p/Xpo1p. Finally, the yeast two-hybrid assay of Fig. 4E shows that MA, but not MA-M4, interacts (directly or through an intermediate) with Crm1p.

These results suggested that in infected cells the MA NES functions to counteract the MA NLS. We predicted that the M4 mutant phenotype would be relieved by a second mutation in the MA NLS. Figure 5 shows that the cellular localization of such a Pr55 double mutant (Pr55M4/16; Fig. 5A) was similar to that of wild-type Pr55, and that the infectivity (Fig. 5C) and virion RNA content (Fig. 5D) of the HXB2-M4/M16 double mutant was nearly normal.

To examine whether regions of Pr55 in addition to MA were involved in viral RNA mislocalization, we performed a series of co-transfection experiments with the wild-type infectious molecular clone, HXB2-WT, and various MA and Pr55 expression constructs. Figure 6A shows that HXB2 genomic RNA was, as expected,

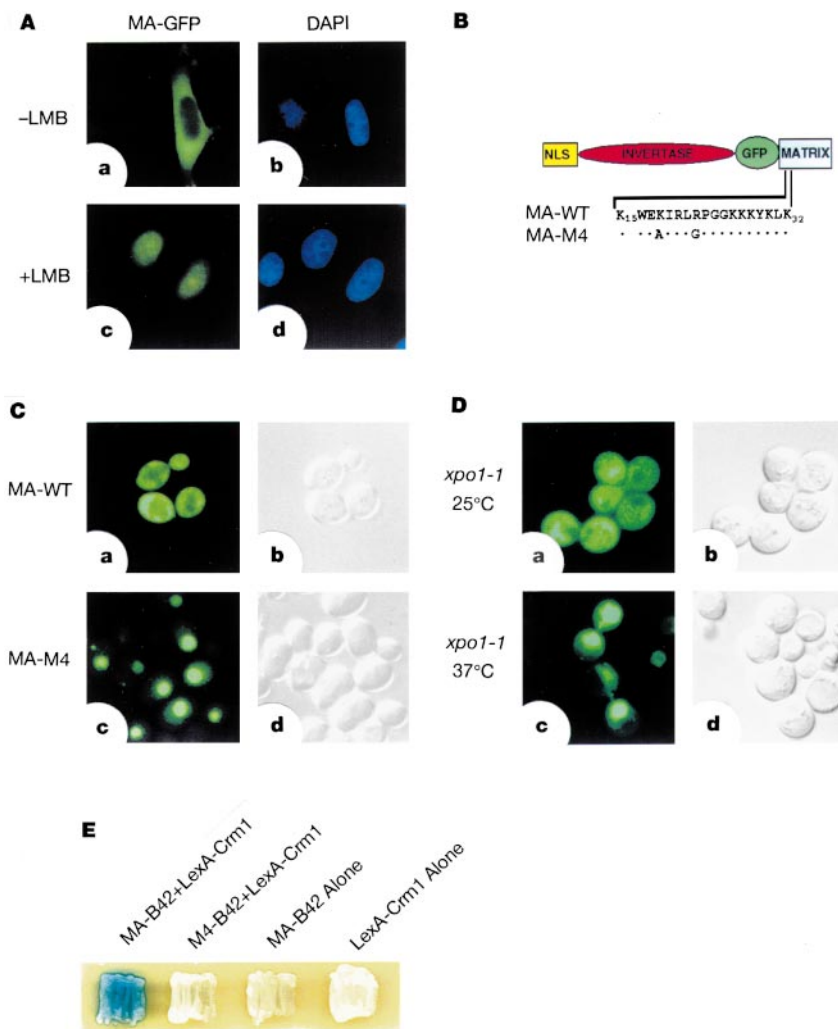


Figure 4 Nuclear export activity of HIV-1 MA. **A**, Cellular distribution of MA is LMB-sensitive. HeLa cells were transiently transfected with a plasmid containing GFP-tagged MA (MA-GFP). At 16 h post-transfection, cells were treated with LMB (0.2 ng ml⁻¹). Nuclei were stained with DAPI. **B**, Schematic diagram of the yeast nuclear export reporter. GFP was fused to yeast invertase flanked by an SV40 Large-T NLS and MA (WT or M4 variant).

C, D, Cellular distribution of MA reporter constructs in wild-type (**C**) and in a mutant yeast strain, *xpo1-1* (**D**). Cells expressing each construct were spheroplasted and viewed by fluorescence (panels **a, c**) or phase-contrast (panels **b, d**) microscopy. **E**, Interaction between HIV-1 MA and Crm1p in a yeast two-hybrid assay. Bait and prey constructs are indicated above each yeast patch.

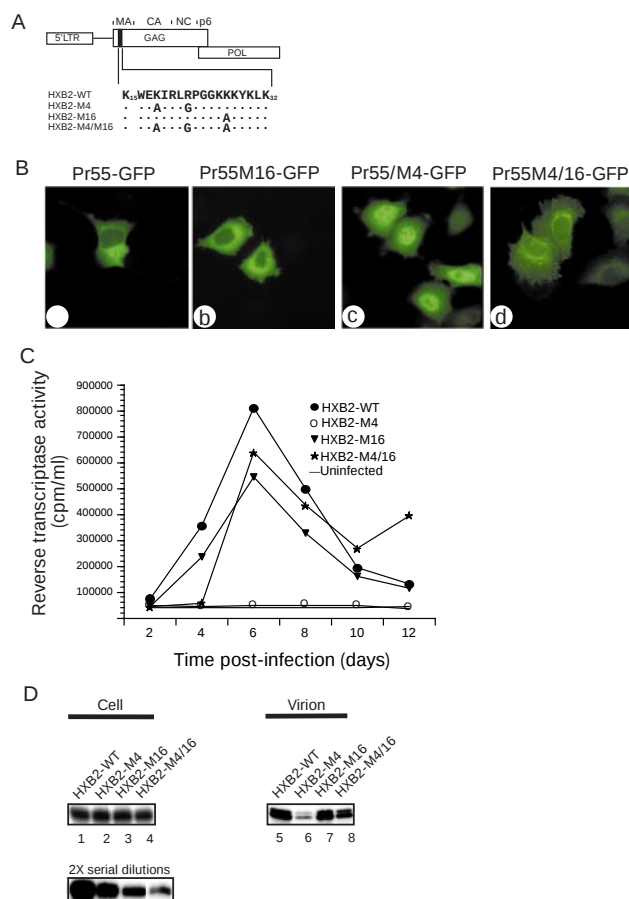


Figure 5 Analysis of a MA NES-NLS double mutant. **A**, Schematic diagram of MA NLS and NES mutants. **B**, Cellular localization of Pr55M16-GFP and Pr55M4/16-GFP. Cellular localization of GFP (panel **a**), Pr55M16-GFP (panel **b**) and Pr55M4/16-GFP (panel **c**) was determined by immunofluorescence microscopy. **C**, An NLS mutation restores the replication defect of HXB2-M4. Replication of wild-type and mutant virus in MT-4 cells was monitored by reverse transcriptase activity in culture supernatants at 2-day intervals ($n = 3$). **D**, RNase protection assay. 10 μ g of total cellular RNA (left) and virion-associated RNA (right) obtained from virus particles (350 ng p24 equivalents) were analysed.

normally cytoplasmic and after co-transfection of Pr55 that distribution was maintained. However, co-transfection of Pr55-M4 resulted in nuclear accumulation of wild-type viral genomic RNA. Also, a minimal MA-M4 (panel **c**) and a Pr55-M4/ Δ NC (panel **d**) derivative failed to mislocalize viral genomic RNA. Thus, whereas the MA-M4 mutation is sufficient to mislocalize Pr55 to the nucleus, the sequence-specific RNA binding activity of nucleocapsid (NC) is also required for viral RNA mislocalization. These results also reveal that Pr55-M4 can mislocalize viral RNA even in the presence of wild-type Pr55, indicating that Pr55-M4 is a dominant-inhibitory mutant.

Finally, we investigated whether Pr55-M4 could mislocalize a non-viral RNA. We inserted a 241-bp viral sequence corresponding to the NC-binding (Psi) site¹⁹ into a β -galactosidase reporter construct to find whether Pr55-M4 would mislocalize this β -gal/Psi RNA to the nucleus. Figure 6B shows that the cellular localization of the β -gal/Psi was cytoplasmic upon co-transfection with Pr55 and predominantly nuclear with Pr55-M4.

Thus we have discovered that MA contains a novel NES and confirmed that MA also contains an NLS, a controversial issue^{4,5,20-22}. On the basis of these results, we propose the following model for the roles of opposing MA cellular localization functions

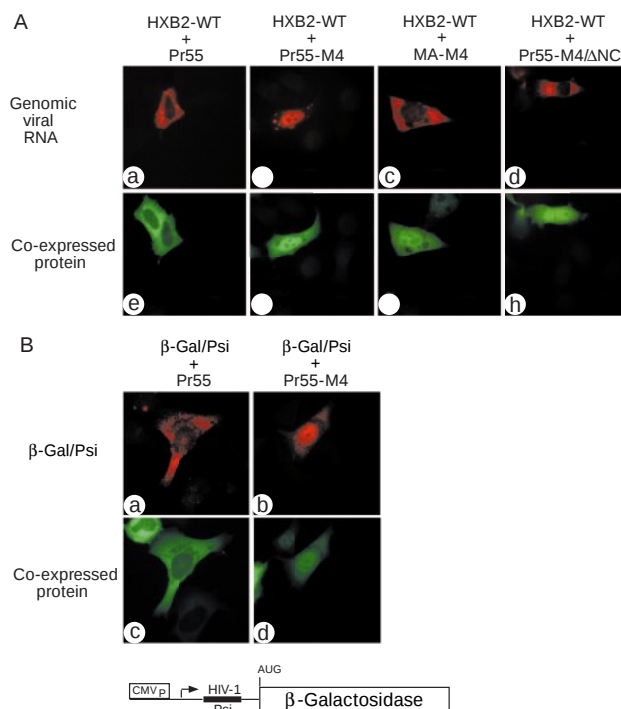


Figure 6 Mislocalization of RNA by Pr55-M4 requires the nucleocapsid-Psi site interaction. **A**, HIV-1 Genomic RNA. HeLa cells were co-transfected with a wild-type HXB2 infectious molecular clone together with plasmids expressing wild-type or mutant MA or Pr55 derivatives. Cellular distribution of wild-type genomic RNA (panels **a-d**) and co-expressed protein (panels **e-h**) are shown. **B**, β -galactosidase/Psi site chimeric RNA. A 241-bp viral sequence encompassing the HIV-1 Psi element was inserted between the transcription start and translation initiation sites of a β -galactosidase (β -gal) reporter construct. Cellular distributions of the β -gal/Psi RNA (panels **a, b**) and co-expressed proteins (panels **c, d**) are shown.

during HIV-1 infection. Following virus entry, the MA NLS participates in targeting the preintegration complex (PIC) to the nucleus. During virus production, the MA NES is the dominant signal and counteracts the NLS to maintain Pr55 in the cytoplasm, the site of virion assembly. An implication of our model is that during virus entry and nuclear localization of the PIC, MA NES is repressed, which could occur, for example, by masking of the NES by a PIC component or inactivation of the NES by phosphorylation. □

Methods

Molecular clones

The $^{18}K \rightarrow A/^{22}R \rightarrow G$ matrix mutation was introduced into HIV-1_{HXB2} (ref. 7) by polymerase chain reaction (PCR)-mediated site-directed mutagenesis and the amplified region (*Bss*H2-*Sph*1) was sequenced to ensure that no other mutation was introduced. MA and Pr55-GFP constructs were separately inserted into a pcDNA3 (Clontech)-based eukaryotic expression vector.

Cells

MT-4 cells were grown in RPMI + 10% fetal bovine serum (FBS). 293T and HeLa cells were grown in DMEM + 10% FBS. All media were supplemented with 10% heat-inactivated fetal bovine serum. For generation of virus particles, 293T cells were transfected with proviral clones (30 μ g) by calcium phosphate/DNA co-precipitation and virions were obtained from cell supernatants 48 s post-transfection.

Detection of viral proteins

HIV-1 gag p24 was measured by enzyme-linked immunosorbent assay (ELISA; DuPont/Immunotech). HIV-1 gp120 was measured by ELISA (ImmunoDiagnostics). Monoclonal antibodies against MA and CA were obtained from Advanced Biotechnologies.

Isolation and detection of viral RNA

Cellular and virion-associated RNA were isolated with Tri-reagent²³ (Molecular Research

Center). RNase protection was performed as recommended by the manufacturer (Ambion). Virion-associated RNA used to assess dimerization status was obtained by proteinase K treatment of virions and phenol extraction of purified virus, as described elsewhere²⁴.

In situ hybridization

Probes were labelled with Cy3 or FITC chromofluors. Cells were fixed with 4% para-formaldehyde and treated as described²⁵.

Yeast strains and transformation

Wild-type cells containing the reporter constructs were grown for 12–18 h at 30 °C in uracil dropout media with 2% sucrose, and then grown for 6 h in uracil dropout media containing 2% galactose. Cells were spheroplasted and viewed by fluorescence microscopy. *Xpo1-1* cells were grown in uracil dropout media with 2% glucose at room temperature and then shifted to 37 °C for 1 h.

Two-hybrid interaction assay

The CRM1 bait construct was provided by M. Rosbash and consists of the whole coding region of CRM1 inserted into pEG202 + PL. Prey constructs were obtained by insertion of wild-type and mutant HIV-1 MA into pJG 4-5. The two-hybrid interaction assay was performed as described²⁶.

Received 14 September; accepted 1 October 1999.

- Lewis, P., Hensel, M. & Emerman, M. Human immunodeficiency virus infection of cells arrested in the cell cycle. *EMBO J.* **11**, 3053–3058 (1992).
- Weinberg, J. B., Matthews, T. J., Cullen, B. R. & Malim, M. H. Productive human immunodeficiency virus type 1 (HIV-1) infection of nonproliferating human monocytes. *J. Exp. Med.* **174**, 1477–1482 (1991).
- Bukrinsky, M. I. *et al.* Active nuclear import of human immunodeficiency virus type 1 preintegration complexes. *Proc. Natl Acad. Sci. USA* **89**, 6580–6584 (1992).
- Bukrinsky, M. I. *et al.* A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells. *Nature* **365**, 666–669 (1993).
- Bukrinsky, M. I. *et al.* Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type 1 with viral nucleic acids following acute infection. *Proc. Natl Acad. Sci. USA* **90**, 6125–6129 (1993).
- Krausslich, H. G. & Welker, R. Intracellular transport of retroviral capsid components. *Curr. Top. Microbiol. Immunol.* **214**, 25–63 (1996).
- Fisher, A. G. *et al.* The trans-activator gene of HTLV-III is essential for virus replication. *Nature* **320**, 367–371 (1986).
- Freed, E. O. HIV-1 gag proteins: diverse functions in the virus life cycle. *Virology* **251**, 1–15 (1998).
- Clever, J. L. & Parslow, T. G. Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *J. Virol.* **71**, 3407–3414 (1997).
- Paillart, J. C. *et al.* A dual role of the putative RNA dimerization initiation site of human immunodeficiency virus type 1 in genomic RNA packaging and proviral DNA synthesis. *J. Virol.* **70**, 8348–8354 (1996).
- Matthews, S. *et al.* Structural similarity between the p17 matrix protein of HIV-1 and interferon-gamma. *Nature* **370**, 666–668 (1994).
- Massiah, M. A. *et al.* Three-dimensional structure of the human immunodeficiency virus type 1 matrix protein. *J. Mol. Biol.* **244**, 198–223 (1994).
- Wolff, B., Sanglier, J. J. & Wang, Y. Leptomycin B is an inhibitor of nuclear export: inhibition of nucleocytoplasmic translocation of the human immunodeficiency virus type 1 (HIV-1) Rev protein and Rev-dependent mRNA. *Chem. Biol.* **4**, 139–147 (1997).
- Nishi, K. *et al.* Leptomycin B targets a regulatory cascade of crm1, a fission yeast nuclear protein, involved in control of higher order chromosome structure and gene expression. *J. Biol. Chem.* **269**, 6320–6324 (1994).
- Fornerod, M., Ohno, M., Yoshida, M. & Mattaj, J. W. CRM1 is an export receptor for leucine-rich nuclear export signals. *Cells* **90**, 1051–1060 (1997).
- Fukuda, M. *et al.* CRM1 is responsible for intracellular transport mediated by the nuclear export signal. *Nature* **390**, 308–311 (1997).
- Ossareh-Nazari, B., Bachelier, F. & Dargemont, C. Evidence for a role of CRM1 in signal-mediated nuclear protein export. *Science* **278**, 141–144 (1997).
- Stade, K., Ford, C. S., Guthrie, C. & Weis, K. Exportin 1 (Crm1p) is an essential nuclear export factor. *Cell* **90**, 1041–1050 (1997).
- Berkowitz, R., Fisher, J. & Goff, S. P. RNA Packaging. *Curr. Top. Microbiol. Immunol.* **214**, 177–218 (1996).
- von Schwedler, U., Kornbluth, R. S. & Trono, D. The nuclear localization signal of the matrix protein of human immunodeficiency virus type 1 allows the establishment of infection in macrophages and quiescent T lymphocytes. *Proc. Natl Acad. Sci. USA* **91**, 6992–6996 (1994).
- Fouchier, R. A., Meyer, B. E., Simon, J. H., Fischer, U. & Malim, M. H. HIV-1 infection of non-dividing cells: evidence that the amino-terminal basic region of the viral matrix protein is important for Gag processing but not for post-entry nuclear import. *EMBO J.* **16**, 4531–4539 (1997).
- Freed, E. O., Englund, G. & Martin, M. A. Role of the basic domain of human immunodeficiency virus type 1 matrix in macrophage infection. *J. Virol.* **69**, 3949–3954 (1995).
- Chomczynski, P. A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. *BioTechniques* **15**, 532–534 (1993).
- Sakuragi, J. I. & Panganiban, A. T. Human immunodeficiency virus type 1 RNA outside the primary encapsidation and dimer linkage region affects RNA dimer stability in vivo. *J. Virol.* **71**, 3250–3254 (1997).
- Kislauskis, E. H., Zhu, X. & Singer, R. H. Sequences responsible for intracellular localization of beta-actin messenger RNA also affect cell phenotype. *J. Cell Biol.* **127**, 441–451 (1994).
- Neville, M., Stutz, F., Lee, L., Davis, L. I. & Rosbash, M. The importin-beta family member crm1p bridges the interaction between Rev and the nuclear pore complex during nuclear export. *Curr. Biol.* **7**, 767–775 (1997).

Acknowledgements

We thank J. Mondor, N. Bakker and L. Ross for manuscript preparation. We also thank M. Yoshida for generous donation of Leptomycin B, and M. Rosbash for yeast strains. We thank J. Kan, J. Teodoro, M. Sharkey, J.-M. Jacque, B. Brichacek, S. Swingler and M. Zapp for scientific discussions. This work was supported in part by grants from the NIH to M.R.G. and M.S., and A.B. is supported in part by a Fogarty International Research Collaboration Award. M.R.G. is an investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to M.R.G. (e-mail: michael.green@umassmed.edu).

Mitochondrial glutamate acts as a messenger in glucose-induced insulin exocytosis

Pierre Maechler & Claes B. Wollheim

Division of Clinical Biochemistry, Department of Internal Medicine, University Medical Centre, 1211 Geneva 4, Switzerland

The hormone insulin is stored in secretory granules and released from the pancreatic β -cells by exocytosis¹. In the consensus model of glucose-stimulated insulin secretion, ATP is generated by mitochondrial metabolism, promoting closure of ATP-sensitive potassium (K_{ATP}) channels, which depolarizes the plasma membrane^{2,3}. Subsequently, opening of voltage-sensitive Ca^{2+} channels increases the cytosolic Ca^{2+} concentration ($[Ca^{2+}]_c$) which constitutes the main trigger initiating insulin exocytosis^{1,3,4}. Nevertheless, the Ca^{2+} signal alone is not sufficient for sustained secretion. Furthermore, glucose elicits a secretory response under conditions of clamped, elevated $[Ca^{2+}]_c$ (refs 5, 6). A mitochondrial messenger must therefore exist which is distinct from ATP^{7,8}. We have now identified this as glutamate. We show that glucose generates glutamate from β -cell mitochondria. A membrane-permeant glutamate analogue sensitizes the glucose-evoked secretory response, acting downstream of mitochondrial metabolism. In permeabilized cells, under conditions of fixed $[Ca^{2+}]_c$, added glutamate directly stimulates insulin exocytosis, independently of mitochondrial function. Glutamate uptake by the secretory granules is likely to be involved, as inhibitors of vesicular glutamate transport suppress the glutamate-evoked exocytosis. These results demonstrate that glutamate acts as an intracellular messenger that couples glucose metabolism to insulin secretion.

In the β -cell, mitochondria have a pivotal role in the regulation of glucose-induced insulin secretion, and mitochondrial activation is required for normal signal transduction^{7–10}. Glucose provides substrates and elevates the Ca^{2+} concentration in the mitochondrial matrix⁹ ($[Ca^{2+}]_m$), resulting in further activation of the tricarboxylic-acid (TCA) cycle¹⁰. We have shown previously that mitochondrial activation generates an unidentified mitochondrial factor, distinct from ATP, which directly triggers insulin exocytosis⁷. We screened molecules derived from mitochondrial metabolism for their secretagogue activity in permeabilized β -cells as described previously⁷. This screening suggested glutamate as a candidate for the putative messenger. Glutamate is formed in the mitochondria from α -ketoglutarate, a TCA-cycle intermediate, by glutamate dehydrogenase¹¹.

When we incubated rat insulinoma INS-1 cells in the presence of a stimulatory glucose concentration (12.8 mM), cellular glutamate levels increased 4.8-fold within 30 min (Fig. 1a). The stimulated values correspond to a total intracellular glutamate concentration of

approximately 0.22 mM, assuming 200 μg protein per μl cell water (corresponding to 10^6 INS-1 cells). In isolated human pancreatic islets¹², glucose (16.7 mM) also augmented glutamate levels from 0.78 ± 0.65 to 3.93 ± 0.04 nmol per mg protein ($P < 0.025$, $n = 4$) during 30 min incubation. The mitochondrial poison FCCP (carbonyl cyanide *p*-trifluoromethoxyphenylhydrazone) inhibited the production of glutamate during glucose stimulation of INS-1 cells ($\sim 43\%$ of cellular glutamate; $n = 5$, $P < 0.01$).

Next we investigated the effects on insulin secretion of a glutamate precursor that was permeable to the cell membrane. Dimethylglutamate (5 mM) caused a leftward shift in the concentration dependency of glucose-stimulated insulin secretion in INS-1 cells (Fig. 1b), in good agreement with observations in rat pancreatic islets¹³. Insulin secretion was not stimulated by dimethylglutamate at basal glucose concentrations or enhanced at optimal glucose concentrations, while secretion was strongly potentiated at intermediate sugar levels. Glutamate is thus formed by mitochondrial metabolism during glucose stimulation (Fig. 1a) and participates in the secretory response (Fig. 1b). These results are consistent with a role for glutamate, derived from glucose, acting downstream of the mitochondria on insulin exocytosis. Accordingly, glutamate dehydrogenase favours glutamate formation rather than providing substrate for the TCA cycle^{11,13} during glucose stimulation of the β -cell. This notion is substantiated by the lack of effect of dimethylglutamate on mitochondrial membrane potential ($\Delta\Psi_m$) (reflecting activation of the respiratory chain¹⁰), measured as the quenching of rhodamine-123 fluorescence (Fig. 2a). Furthermore, dimethylglutamate did not increase $[\text{Ca}^{2+}]_m$ in INS-1 cells expressing the

Ca^{2+} -sensitive photoprotein aequorin in the mitochondria⁹ (Fig. 2b). By contrast, stimulation with methyl succinate, a cell-membrane-permeant precursor for succinate, a TCA cycle intermediate⁷, confirms that changes in $\Delta\Psi_m$ and $[\text{Ca}^{2+}]_m$ indeed reflect activation of mitochondrial respiration¹⁰ (Fig. 2a, b). Glucose and methyl succinate, both insulin secretagogues, initially increased $[\text{Ca}^{2+}]_c$ and hyperpolarized $\Delta\Psi_m$, resulting in a large increase in $[\text{Ca}^{2+}]_m$ (refs 7, 9, 14). Therefore, glucose and methyl succinate stimulate insulin secretion by activating mitochondrial metabolism^{7,9,14}, whereas the metabolic product glutamate apparently participates directly in the secretory process downstream of mitochondrial activation.

To investigate the putative messenger function of glutamate further, we used INS-1 cells permeabilized with *Staphylococcus* α -toxin; in this preparation, insulin secretion in response to mitochondrial substrates is preserved at permissive $[\text{Ca}^{2+}]_c$ (ref. 7). Succinate (1 mM), but not glutamate (1 mM), hyperpolarized the $\Delta\Psi_m$ (Fig. 2c). This hyperpolarization facilitated Ca^{2+} uptake by the mitochondria above a threshold of $[\text{Ca}^{2+}]_c$ (≈ 300 nM)¹⁰. At clamped $[\text{Ca}^{2+}]_c$ (500 nM), succinate induced a rise in $[\text{Ca}^{2+}]_m$ (Fig. 2d, thick line). This was not observed at 100 nM $[\text{Ca}^{2+}]_c$ (Fig. 2d, thin line), a concentration corresponding to basal $[\text{Ca}^{2+}]_c$ in intact cells⁹. In contrast to succinate, glutamate did not activate the mitochondria as monitored by $\Delta\Psi_m$ and $[\text{Ca}^{2+}]_m$ (Fig. 2c, d).

We next measured insulin secretion in the effluent of permeabilized cells perfused with 1 mM ATP and 500 nM $[\text{Ca}^{2+}]_c$. Insulin secretion was stimulated by 1 mM succinate (Fig. 3a), which activated the mitochondrial metabolism by providing substrate and increasing $[\text{Ca}^{2+}]_m$ (Fig. 2d). Despite the absence of a rise in $[\text{Ca}^{2+}]_m$, glutamate (1 mM) increased insulin secretion from 0.47 ± 0.08 to 1.09 ± 0.25 ng per min ($P < 0.05$, $n = 7$) (Fig. 3b), that is, to the same extent as succinate. At 0.1 mM, glutamate still slightly augmented insulin release, from 0.53 ± 0.03 to 0.84 ± 0.13 ng per min ($P < 0.05$, $n = 4$), but 0.02 mM glutamate was ineffective (data not shown). This concentration range agrees well with the glutamate levels we found in glucose-stimulated INS-1 cells (0.22 mM following 30 min incubation; see above).

Succinate-induced insulin secretion requires the associated rise in $[\text{Ca}^{2+}]_m$ for the generation of a messenger^{7,8}. We propose that the messenger is glutamate because this molecule was able to trigger exocytosis without mitochondrial activation at 500 nM $[\text{Ca}^{2+}]_c$ (Fig. 3b). Nevertheless, an elevated Ca^{2+} concentration in the cytosol

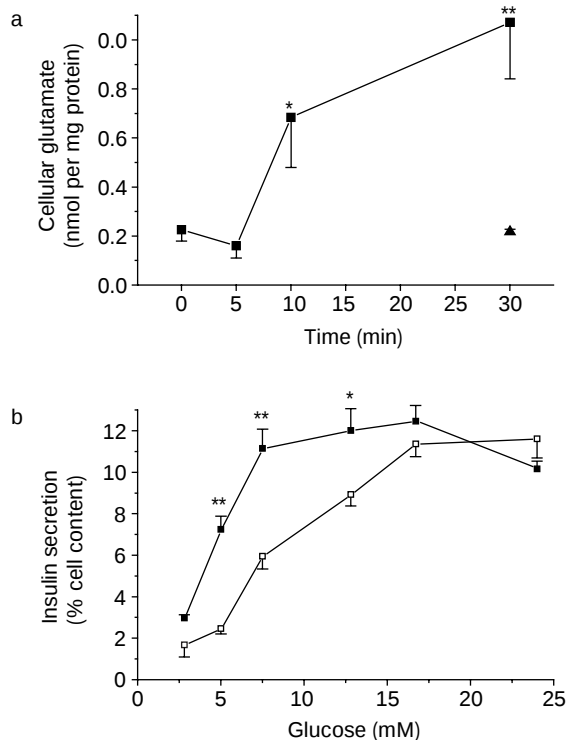


Figure 1 Glutamate production and effects of dimethylglutamate in INS-1 insulinoma cells. **a**, Changes in cellular glutamate levels during incubation of INS-1 cells in the presence of 2.8 mM (filled triangles) or 12.8 mM (filled squares) glucose. The results are means $\pm$ s.e.m., $n = 6$ independent experiments done in duplicate; $*P < 0.05$, $**P < 0.005$ versus $t = 0$. **b**, Dimethylglutamate (5 mM met-Glut, filled squares) sensitizes glucose-induced insulin secretion in INS-1 cells. Non-esterified, cell-impermeant glutamate does not enhance insulin secretion (data not shown). The results are means $\pm$ s.e.m., $n = 4$ of one out of three similar experiments; $*P < 0.05$, $**P < 0.005$ versus corresponding glucose control (open squares).

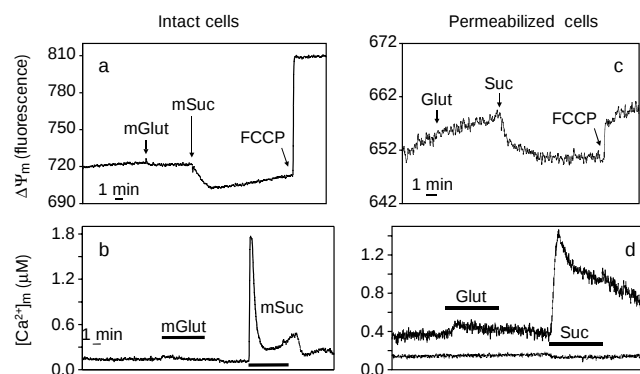


Figure 2 Mitochondrial membrane potential ($\Delta\Psi_m$) and $[\text{Ca}^{2+}]_m$ in intact and permeabilized INS-1 cells. **a**, **b**, Effects of cell-permeant dimethylester-L-glutamate (5 mM, mGlut) and monomethylester succinate (5 mM, mSuc) on $\Delta\Psi_m$ (**a**) and $[\text{Ca}^{2+}]_m$ (**b**). **c**, **d**, Effects of 1 mM L-glutamate (Glut) or 1 mM succinate (Suc) on $\Delta\Psi_m$ (**c**) and $[\text{Ca}^{2+}]_m$ (**d**) in α -toxin-permeabilized INS-1 cells incubated at 500 nM $[\text{Ca}^{2+}]_c$ (thin bottom line in **d** at 100 nM $[\text{Ca}^{2+}]_c$). In **a** and **c**, $\Delta\Psi_m$ is measured in a cuvette and mitochondria are finally maximally depolarized (upward deflection) with 1 μM of the protonophore FCCP. The traces are representative of 3–5 independent experiments.

remains a prerequisite for glutamate-induced activation of the exocytotic machinery. Accordingly, at basal 100 nM $[Ca^{2+}]_c$, glutamate failed to stimulate insulin secretion (Fig. 3c). As insulin exocytosis requires not only Ca^{2+} but also ATP¹, we next investigated the interactions of ATP and glutamate. Even when the ambient ATP was increased 10-fold to the saturating concentration of 10 mM (at 500 nM $[Ca^{2+}]_c$), insulin secretion (% cell content per 10 min) was stimulated from 5.29 ± 0.57 to 10.26 ± 2.29 by 1 mM glutamate ($P < 0.05$) compared with 11.58 ± 2.08 by 1 mM succinate ($P < 0.02$). Thus, ATP does not mediate the effect of glutamate on exocytosis.

We also monitored the changes in cytosolic [ATP] together with insulin secretion in the same permeabilized cells expressing the ATP sensor luciferase in the cytosol¹⁵. Activation of the mitochondrial respiratory chain, seen as hyperpolarization of $\Delta\Psi_m$, promotes the synthesis of ATP in mitochondria. Succinate (1 mM), which hyperpolarized the $\Delta\Psi_m$ (Fig. 2c), generated ATP (2.4-fold) in permeabilized

cells perfused with basal 1 mM ATP and 500 nM $[Ca^{2+}]_c$ (Fig. 3d). Unexpectedly, 1 mM glutamate also increased cytosolic [ATP] (1.9-fold) (Fig. 3d), without activating the respiratory chain, as demonstrated by the lack of effect on $\Delta\Psi_m$ (Fig. 2c).

Monitoring of [ATP] in perfused permeabilized cells reflects the net cytosolic ATP concentration resulting from exogenous (buffer) and endogenous (mitochondria) ATP supplies balanced by ATP consumption owing to ATPase activity (mainly Ca^{2+} -ATPases). Inhibition of the mitochondrial ATP synthase¹⁰ with oligomycin lowered basal cytosolic [ATP] but did not abolish the effect of glutamate (Fig. 4b). Moreover, the effect on insulin secretion was well preserved (Fig. 4a). Under these conditions, succinate did not increase cytosolic [ATP] or insulin exocytosis (data not shown). These results establish that the secretory response evoked by the addition of glutamate does not require mitochondrial metabolism and suggest a non-mitochondrial origin for the ATP augmented by glutamate.

Like chromaffin granules¹⁶ and synaptic vesicles¹⁷, insulin-containing granules^{18,19} are acidic inside (pH ≈ 5.0). The pH gradient is generated by a proton pump (vacuolar-type H^+ -ATPase) using cytosolic ATP^{16,18}. As a consequence, a membrane potential is created which is positive inside^{17,18,20}. The sum of the pH gradient and the membrane potential, the electrochemical proton gradient,

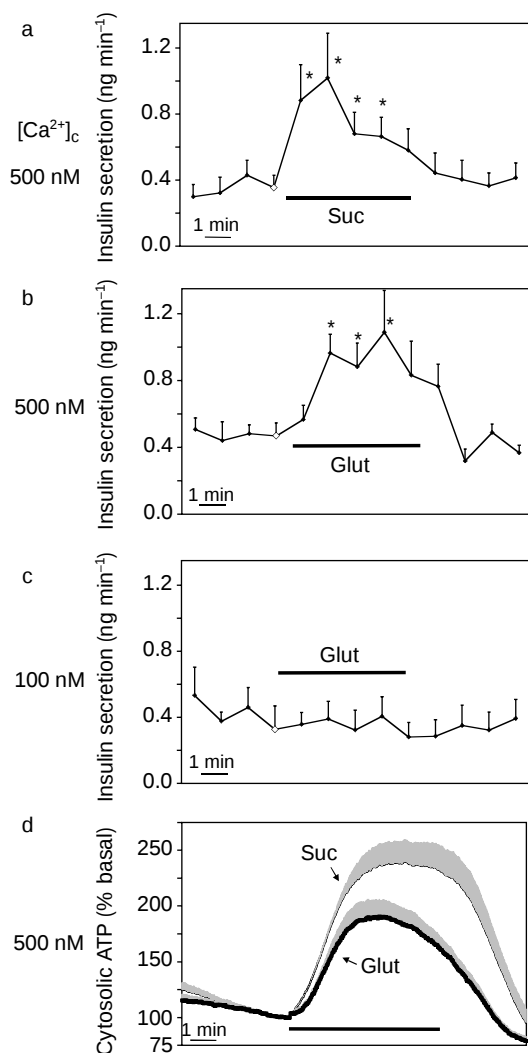


Figure 3 Effects of succinate and glutamate on insulin secretion and cytosolic [ATP] in α -toxin-permeabilized INS-1 cells. Cells are perfused at 500 nM (a, b, d) or 100 nM (c) free Ca^{2+} and 1 mM ATP. Insulin secretion is measured (a–c) in the effluent of cell preparations monitored for cytosolic [ATP] (d) through bioluminescence of cytosolic luciferase. Effects of: a, d (thin line), 1 mM succinate (Suc); b, d (thick line), 1 mM L-glutamate (Glut) at 500 nM $[Ca^{2+}]_c$; c, 1 mM Glut at 100 nM $[Ca^{2+}]_c$. The results are means $\pm$ s.e.m. (depicted by a shaded area for ATP), $n = 6$ (a, d), 7 (b, d) and 4 (c) independent experiments; in a and b, $*P < 0.05$ versus the time point just preceding stimulation (open diamond).

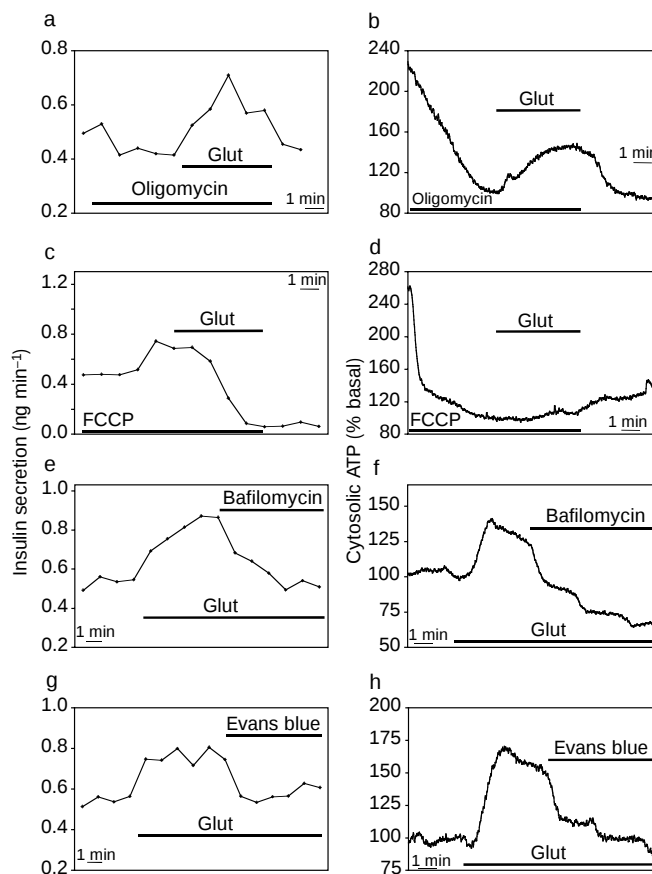


Figure 4 Effects of different inhibitors on the action of glutamate on insulin secretion (left) and cytosolic [ATP] (right) in permeabilized INS-1 cells. Cells are perfused at 500 nM Ca^{2+} and 1 mM ATP. Insulin secretion is measured in the effluent of the cell preparations monitored for cytosolic [ATP]. a, b, Effects of 1 mM L-glutamate (Glut) in the presence of $1 \mu g \text{ ml}^{-1}$ of the mitochondrial ATP synthase inhibitor oligomycin. c, d, Effects of 1 mM Glut in the presence of $1 \mu M$ of the protonophore FCCP. e, f, Effects of 100 nM of the vacuolar-type H^+ -ATPase inhibitor bafilomycin during stimulation with 1 mM Glut. g, h, Effects of $2 \mu M$ of the glutamate transport inhibitor Evans blue during stimulation with 1 mM Glut. The traces are representative of 3–5 independent experiments.

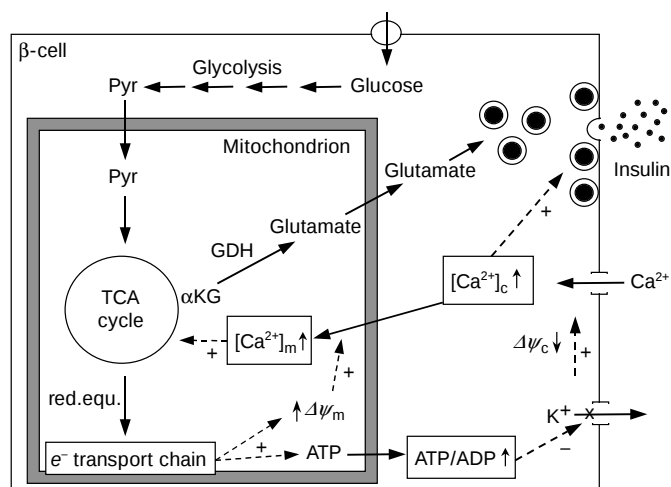


Figure 5 Proposed model for coupling glucose metabolism to insulin secretion in the β -cell. Glycolysis converts glucose to pyruvate (Pyr), which enters the mitochondrion and the TCA cycle resulting in the transfer of reducing equivalents (red. equ.) to the electron (e^-) transport chain, hyperpolarization of $\Delta\Psi_c$ and generation of ATP. Subsequently, closure of K_{ATP} channels depolarizes $\Delta\Psi_c$ which opens voltage-sensitive Ca^{2+} channels, raising $[Ca^{2+}]_c$ and triggering insulin exocytosis. At this time, under conditions of elevated $[Ca^{2+}]_c$, hyperpolarized $\Delta\Psi_m$ increases $[Ca^{2+}]_m$, further activating the TCA cycle. Glutamate is then formed from α -ketoglutarate (α KG) by the glutamate dehydrogenase (GDH). Glutamate uptake by granules leads to the second phase of insulin secretion.

drives glutamate uptake into the vesicular compartment via a specific transporter^{17,21}. Dissipation of both the proton gradient and the vesicular membrane potential with protonophores²⁰ inhibits glutamate uptake^{17,21}. Either FCCP or oligomycin can deplete ATP, but only FCCP affects the vesicular proton gradient, which is left intact by oligomycin. In the presence of FCCP, insulin release was transiently augmented, probably through a direct effect on the granules²¹ and, thereafter, glutamate no longer increased insulin secretion and [ATP] (Fig. 4c, d). This indicates that glutamate uptake by the secretory granules is implicated in the elevation of cytosolic [ATP]. This could be explained by two mechanisms. First, glutamate acidifies vesicles^{17,21} and may thereby reduce the consumption of ATP by the vacuolar-type H^+ -ATPase. Second, it is possible that glutamate uptake promotes ATP production by reversal of the vacuolar-type H^+ -ATPase^{16,22}. The effect of glutamate on cytosolic [ATP] was abolished by specific inhibition of the vacuolar ATPase (Fig. 4f) or the vesicular glutamate transporter (Fig. 4h), further indicating the granules as the ATP source.

Bafilomycin (which blocks the vacuolar ATPase) and FCCP inhibit glutamate uptake by synaptic vesicles²³. The abrogation of glutamate-induced insulin exocytosis observed with FCCP and bafilomycin (Fig. 4c, e) is probably secondary to blockage of glutamate uptake by the secretory granules. This notion was further substantiated by results obtained with Evans blue, a competitive inhibitor of the vesicular glutamate transporter^{21,23}. Evans blue abolished the glutamate-induced insulin secretion (Fig. 4g), without affecting insulin release at basal conditions (500 nM Ca^{2+}) (not shown).

In the brain, glutamate is the primary extracellular messenger stored in synaptic vesicles²¹. We propose a novel role for glutamate as an intracellular messenger in insulin exocytosis. Glutamate uptake would render the insulin granules secretion competent. This could occur through the reduction of granular membrane potential¹⁷ or, possibly, through Ca^{2+} uptake into the granules, as Ca^{2+} depletion of this compartment inhibits insulin exocytosis²⁴. It is of interest in this context that Ca^{2+} -evoked exocytosis in

permeabilized mast cells is strictly dependent on the presence of glutamate²⁵. Glutamate uptake could also result in granule swelling, enabling the granule to fuse with the plasma membrane, a phenomenon described in zymogen granules²⁶. The production of glutamate during glucose stimulation unravels an essential role for the mitochondria in exocytosis. Glutamate does not initiate the secretory response and its production from glucose requires Ca^{2+} entry into the cell and a rise in $[Ca^{2+}]_m$, which in turn increases the carbon flux through the TCA cycle^{8,27}. The lag in glutamate generation (5–10 min; Fig. 1a) suggests that the molecule is implicated in the second, sustained phase of the biphasic glucose-induced insulin secretion⁵.

Therefore, glutamate acts as an intracellular messenger in the normal regulation of insulin secretion in response to glucose (Fig. 5). It is noteworthy that children with mutations in the glutamate dehydrogenase gene, which results in excessive enzyme activity, display hypersecretion of insulin²⁸. □

Methods

Cell culture

INS-1 cells were cultured in RPMI 1640 medium with 5% fetal calf serum (FCS) as previously described⁷. Stable clones of INS-1 cells expressing the Ca^{2+} -sensitive photo-protein aequorin targeted to the mitochondria (INS-1/EK3) were used to measure $[Ca^{2+}]_m$ (ref. 9). Clonal INS-1 lines expressing cytosolic luciferase (INS-r3-LUC7) were used for monitoring cytosolic [ATP] in living cells¹⁵. Human pancreatic islets were isolated by collagenase digestion and cultured free floating in CMRL-1066 medium with 10% FCS before the experiments¹².

Cellular glutamate determination

Attached INS-1 cells or free-floating human islets were cultured in 10-cm Petri dishes and preincubated for 2 h in glucose-free RPMI 1640 medium at 37 °C. Cells were then incubated for the indicated times at 37 °C in Krebs–Ringer bicarbonate HEPES buffer (KRBH) containing (in mM): 135 NaCl, 3.6 KCl, 10 HEPES (pH 7.4), 5 $NaHCO_3$, 0.5 NaH_2PO_4 , 0.5 $MgCl_2$, 1.5 $CaCl_2$ and 2.8 glucose. The stimulatory glucose concentration was 12.8 mM for INS-1 cells and 16.7 mM for human islets. Stimulation was ended by putting the Petri dishes on ice and adding 1 ml of lysis buffer (20 mM Tris–HCl pH 8.0, 2 mM CDTA, 0.2% Tween-20) to the cells after discarding the incubation buffer. Following protein determination (Bradford's assay), glutamate levels were measured in cell homogenates by monitoring NADH fluorescence augmentation during oxidation of glutamate in the cell extracts in the presence of an excess of glutamate dehydrogenase (Boehringer-Mannheim) as described²⁹, with modifications. NADH fluorescence, excited at 340 nm, was measured at 460 nm in an LS-50B fluorimeter (Perkin-Elmer) before and after a 30-min incubation at room temperature in 1.5 ml buffer (50 mM Tris–HCl pH 9.5, 2.6 mM EDTA, 1.4 mM NAD, 1 mM ADP) and basal NADH was subtracted. Glutamate levels were corrected according to internal standards.

Cell permeabilization

Attached INS-1 cells were grown on coverslips coated with an extracellular matrix⁸ and permeabilized after a 3–5 day culture period. Cells were first washed with a Ca^{2+} -free KRBH buffer before permeabilization with *Staphylococcus aureus* α -toxin (2.5 μ g per coverslip, that is, per $4\text{--}5 \times 10^5$ cells) at 37 °C for 10 min in 100 μ l of an intracellular-type buffer adjusted to approximately 100 nM free Ca^{2+} (140 mM KCl, 5 mM NaCl, 7 mM $MgSO_4$, 20 mM HEPES, pH 7.0, 1 mM ATP, 10.2 mM EGTA, 1.65 mM $CaCl_2$)⁷. Where mentioned, perfusion was done with the same low Ca^{2+} intracellular buffer or switched to 500 nM free Ca^{2+} concentration (140 mM KCl, 5 mM NaCl, 7 mM $MgSO_4$, 20 mM HEPES, pH 7.0, 10.2 mM EGTA, 6.67 mM $CaCl_2$) with 1 or 10 mM ATP as indicated⁷.

Static insulin secretion

For static incubations, INS-1 cells were cultured for 3–5 days in complete RPMI 1640 medium. Before the experiments, cells were maintained for 2 h in glucose-free culture medium, washed and preincubated in glucose-free KRBH for 30 min and then incubated in the presence of the appropriate stimuli for 30 min at 37 °C. Insulin secretion and cellular insulin content extracted with acid–ethanol were determined by radioimmunoassay using rat insulin as a standard⁷. Static incubations of permeabilized cells were performed as previously described⁷. Where indicated, insulin release was measured in the effluent of permeabilized cells undergoing monitoring of luminescence (see next section). For all insulin secretion experiments, 0.1% of bovine serum albumin (Sigma) was added to buffers as carrier.

Mitochondrial membrane potential

The $\Delta\Psi_m$ was measured in a suspension of intact or α -toxin-permeabilized cells previously loaded with 10 μ g ml^{-1} rhodamine-123. The fluorescence was excited at 490 nm and measured at 530 nm in an LS-50B fluorimeter at 37 °C in a cuvette with gentle stirring⁷.

Measurements of luminescence and insulin secretion

Luciferase- or aequorin-expressing cells were seeded on coverslips 3–5 days before analysis. Before luminescence measurements, cells were maintained in glucose-free RPMI 1640 for 2 h at 37 °C. This period also served to load aequorin-expressing cells with 2.5 μ M coelenterazine, the prosthetic group of aequorin⁹. Luminescence was measured in a thermostatted chamber at 37 °C by a photon detector connected to a photomultiplier apparatus (EMI 9789, Thorn-EMI) and data were collected each second on a computer photon-counting board (EMI C660)⁹. The cells were perfused constantly at a rate of 1 ml min⁻¹ and, where indicated, fractions (one per min) were collected from the effluent for insulin measurement. Intact cells were perfused with KRBH and 10 μ M beetle luciferin was added to the buffer for cytosolic [ATP] monitoring in luciferase-expressing cells¹⁵.

Received 30 April; accepted 9 September 1999.

1. Wollheim, C. B., Lang, J. & Regazzi, R. The exocytotic process of insulin secretion and its regulation by Ca²⁺ and G-proteins. *Diabetes Rev.* **4**, 276–297 (1996).
2. Matschinsky, F. M. A lesson in metabolic regulation inspired by the glucokinase glucose sensor paradigm. *Diabetes* **45**, 223–241 (1996).
3. Ashcroft, F. M. et al. Stimulus-secretion coupling in pancreatic beta cells. *J. Cell Biochem.* **55**, 54–65 (1994).
4. Lang, J. Molecular mechanisms and regulation of insulin exocytosis as a paradigm of endocrine secretion. *Eur. J. Biochem.* **259**, 3–17 (1999).
5. Gembal, M., Gilon, P. & Henquin, J. C. Evidence that glucose can control insulin release independently from its action on ATP-sensitive K⁺ channels in mouse B cells. *J. Clin. Invest.* **89**, 1288–1295 (1992).
6. Sato, Y., Aizawa, T., Komatsu, M., Okada, N. & Yamada, T. Dual functional role of membrane depolarization/Ca²⁺ influx in rat pancreatic B-cell. *Diabetes* **41**, 438–443 (1992).
7. Maechler, P., Kennedy, E. D., Pozzan, T. & Wollheim, C. B. Mitochondrial activation directly triggers the exocytosis of insulin in permeabilized pancreatic β -cells. *EMBO J.* **16**, 3833–3841 (1997).
8. Maechler, P., Kennedy, E. D., Wang, H. & Wollheim, C. B. Desensitization of mitochondrial Ca²⁺ and insulin secretion responses in the beta cell. *J. Biol. Chem.* **273**, 20770–20778 (1998).
9. Kennedy, E. D. et al. Glucose-stimulated insulin secretion correlates with changes in mitochondrial and cytosolic Ca²⁺ in aequorin-expressing INS-1 cells. *J. Clin. Invest.* **98**, 2524–2538 (1996).
10. Duchon, M. R. Contributions of mitochondria to animal physiology: from homeostatic sensor to calcium signalling and cell death. *J. Physiol. (Lond.)* **516**, 1–17 (1999).
11. Fisher, H. F. L-glutamate dehydrogenase from bovine liver. *Meth. Enzymol.* **113**, 16–27 (1985).
12. Janjic, D. et al. Improved insulin secretion of cryopreserved human islets by antioxidant treatment. *Pancreas* **13**, 166–172 (1996).
13. Sener, A. et al. Insulinotropic action of glutamic acid dimethyl ester. *Am. J. Physiol.* **267**, E573–E584 (1994).
14. Eto, K. et al. Role of NADH shuttle system in glucose-induced activation of mitochondrial metabolism and insulin secretion. *Science* **283**, 981–985 (1999).
15. Maechler, P., Wang, H. & Wollheim, C. B. Continuous monitoring of ATP levels in living insulin secreting cells expressing cytosolic firefly luciferase. *FEBS Lett.* **422**, 328–332 (1998).
16. Johnson, R. G. Jr Proton pumps and chemiosmotic coupling as a generalized mechanism for neurotransmitter and hormone transport. *Ann. N. Y. Acad. Sci.* **493**, 162–177 (1987).
17. Maycox, P. R., Deckwerth, T., Hell, J. W. & Jahn, R. Glutamate uptake by brain synaptic vesicles. Energy dependence of transport reconstitution in proteoliposomes. *J. Biol. Chem.* **263**, 15423–15428 (1988).
18. Hutton, J. C. The internal pH and membrane potential of the insulin-secretory granule. *Biochem. J.* **204**, 171–178 (1982).
19. Orci, L. et al. Conversion of proinsulin to insulin occurs coordinately with acidification of maturing secretory vesicles. *J. Cell Biol.* **103**, 2273–2281 (1986).
20. Breckenridge, L. J. & Almers, W. Currents through the fusion pore that forms during exocytosis of a secretory vesicle. *Nature* **328**, 814–817 (1987).
21. Ozkan, E. D. & Ueda, T. Glutamate transport and storage in synaptic vesicles. *Jpn. J. Pharmacol.* **77**, 1–10 (1998).
22. Roisin, M. P., Scherman, D. & Henry, J. P. Synthesis of ATP by an artificially imposed electrochemical proton gradient in chromaffin granule ghosts. *FEBS Lett.* **115**, 143–147 (1980).
23. Roseth, S., Fykse, E. M. & Fonnum, F. Uptake of L-glutamate into rat brain synaptic vesicles: effect of inhibitors that bind specifically to the glutamate transporter. *J. Neurochem.* **65**, 96–103 (1995).
24. Scheenen, W. J., Wollheim, C. B., Pozzan, T. & Fasolato, C. Ca²⁺ depletion from granules inhibits exocytosis. A study with insulin-secreting cells. *J. Biol. Chem.* **273**, 19002–19008 (1998).
25. Churcher, Y. & Gomperts, B. D. ATP-dependent and ATP-independent pathways of exocytosis revealed by interchanging glutamate and chloride as the major anion in permeabilized mast cells. *Cell Regul.* **1**, 337–346 (1990).
26. Jena, B. P. et al. G_i regulation of secretory vesicle swelling examined by atomic force microscopy. *Proc. Natl Acad. Sci. USA* **94**, 13317–13322 (1997).
27. Hyder, F. et al. Increased tricarboxylic acid cycle flux in rat brain during forepaw stimulation detected with ¹H-[¹³C]NMR. *Proc. Natl Acad. Sci. USA* **93**, 7612–7617 (1996).
28. Stanley, C. A. et al. Hyperinsulinism and hyperammonemia in infants with regulatory mutations of the glutamate dehydrogenase gene. *N. Engl. J. Med.* **338**, 1352–1357 (1998).
29. Cho, S. W., Lee, J. & Choi, S. Y. Two soluble forms of glutamate dehydrogenase isoproteins from bovine brain. *Eur. J. Biochem.* **233**, 340–346 (1995).

Acknowledgements

We thank C. Bartley and G. Chaffard for technical assistance, J. Lou, J. Oberholzer and P. Morel (Department of Surgery, University Hospital of Geneva) for supplying human islets, A. Valeva (Institute of Medical Microbiology, University of Mainz) for α -toxin and T. Pozzan, P. Antinuzzi and H. Ishihara for discussions. This study was supported by the Swiss National Science Foundation and the AETAS Foundation (Geneva).

Correspondence and requests for materials should be addressed to P.M. (e-mail: pierre.maechler@medecine.unige.ch).

The TOR signalling pathway controls nuclear localization of nutrient-regulated transcription factors

Thomas Beck & Michael N. Hall

Department of Biochemistry, Biozentrum, University of Basel, CH-4056 Basel, Switzerland

The rapamycin-sensitive TOR signalling pathway in *Saccharomyces cerevisiae* activates a cell-growth program in response to nutrients such as nitrogen and carbon^{1–4}. The TOR1 and TOR2 kinases (TOR) control cytoplasmic protein synthesis and degradation through the conserved TAP42 protein^{5–8}. Upon phosphorylation by TOR, TAP42 binds and possibly inhibits type 2A and type-2A-related phosphatases^{6–8}; however, the mechanism by which TOR controls nuclear events such as global repression of starvation-specific transcription is unknown. Here we show that TOR prevents transcription of genes expressed upon nitrogen limitation by promoting the association of the GATA transcription factor GLN3 with the cytoplasmic protein URE2. The binding of GLN3 to URE2 requires TOR-dependent phosphorylation of GLN3. Phosphorylation and cytoplasmic retention of GLN3 are also dependent on the TOR effector TAP42, and are antagonized by the type-2A-related phosphatase SIT4. TOR inhibits expression of carbon-source-regulated genes by stimulating the binding of the transcriptional activators MSN2 and MSN4 to the cytoplasmic 14-3-3 protein BMH2. Thus, the TOR signalling pathway broadly controls nutrient metabolism by sequestering several transcription factors in the cytoplasm.

Nitrogen limitation activates the partially redundant GATA factors GLN3 and GAT1^{9–11}. We investigated a role of TOR in GLN3- and GAT1-dependent transcription. Wild-type cells were grown in rich medium to early logarithmic phase, treated with

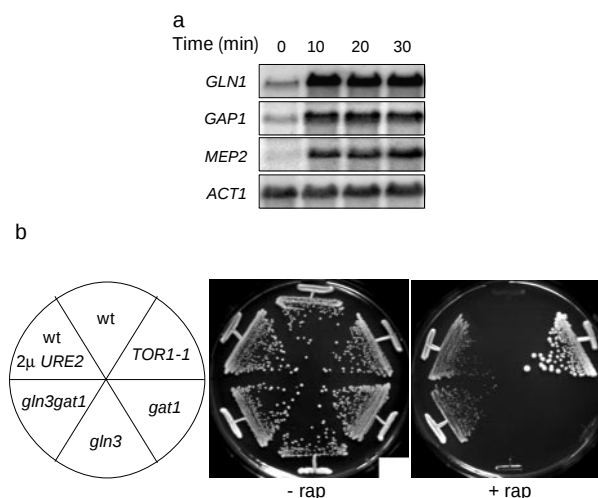


Figure 1 TOR inhibits the GATA transcription factors GLN3 and GAT1. **a**, Induction of GLN3- and GAT1-dependent transcription in response to TOR inactivation by rapamycin treatment. GLN1, GAP1, MEP2 and ACT1 transcripts from cells treated with rapamycin for the times indicated were probed. **b**, Overexpression of URE2 or deletion of GLN3 and GAT1 confers rapamycin resistance. Growth of a wild-type strain overexpressing URE2 (2 μ URE2; pTB376), strains deleted for GLN3 (gln3), GAT1 (gat1) or GLN3 and GAT1 (gln3 gat1), a wild-type (wt) strain and a rapamycin-resistant TOR1-1 strain¹⁴ on rich medium (-rap) and rich medium supplemented with 200 ng ml⁻¹ of rapamycin (+rap) was compared. The -rap and +rap plates were incubated at 30 °C for 2 days and 5 days, respectively.

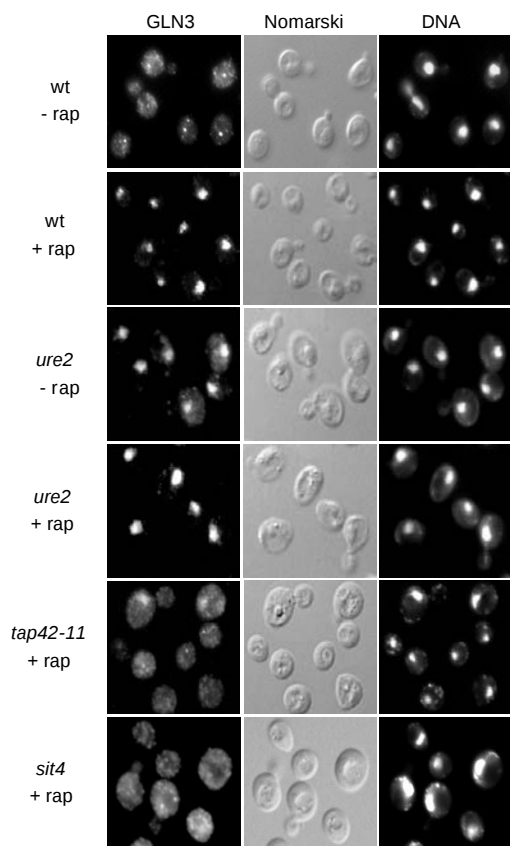


Figure 2 GLN3 localization is controlled by TOR, URE2, TAP42 and SIT4. Localization of GLN3-myc¹³ (GLN3) in a wild type (wt) (TB123) and in *ure2* (TB138-1a), *tap42-11* (TB139-5b) and *sit4* (TB136-2a) mutant cells, grown at 30 °C (*tap42-11* mutant at 24 °C) in rich medium and treated (+rap) or not treated (–rap) with rapamycin. Wild-type and *ure2* cells were treated with rapamycin for 10 minutes. The *tap42-11* and *sit4* cells were treated with rapamycin for 30 minutes. Cells and DNA were visualized by Nomarski optics and DAPI staining (DNA).

rapamycin to inactivate the TOR proteins, and the GLN3- or GAT1-dependent transcripts *MEP2*, *GAP1* and *GLN1* were probed. Within 10 minutes of rapamycin treatment, the levels of these transcripts were induced roughly 10-fold, whereas the *ACT1* mRNA level remained unchanged (Fig. 1a). The induction of at least *MEP2* and *GLN1* was blocked in a *gln3 gat1* double mutant (data not shown). Thus, TOR prevents GLN3- and GAT1-dependent transcription. Because inappropriate expression of GLN3, and possibly GAT1, is toxic¹², we investigated whether the inhibition of growth by rapamycin is, at least in part, due to rapamycin-induced activation of the GATA factors. Strains deleted for *GLN3* and *GAT1* either individually or in combination were monitored for growth on rich medium containing rapamycin (Fig. 1b). Whereas the single deletion strains (*gln3* and *gat1*) were sensitive to rapamycin, the double knockout strain (*gln3 gat1*) was weakly resistant to the drug. The rapamycin resistance of the double mutant was similar to that of a wild-type strain overexpressing URE2, a negative regulator of nitrogen catabolic genes^{9,13} (Fig. 1b). Thus, part of the role of TOR in promoting growth is to inactivate GLN3 and GAT1.

To investigate the mechanism by which TOR inactivates GLN3 and GAT1, we constructed strains bearing genomically tagged versions of *GLN3* (*GLN3-myc¹³*) and *GAT1* (*GAT1-HA³*) and visualized the proteins, by indirect immunofluorescence, in untreated cells and in cells treated with rapamycin for various times. In rapamycin-untreated cells, GLN3 and GAT1 were detected only in the cytoplasm. Rapamycin treatment caused a rapid translocation of GLN3 and GAT1 from the cytoplasm into the nucleus. Within 10 minutes, a pronounced nuclear accumulation of GLN3

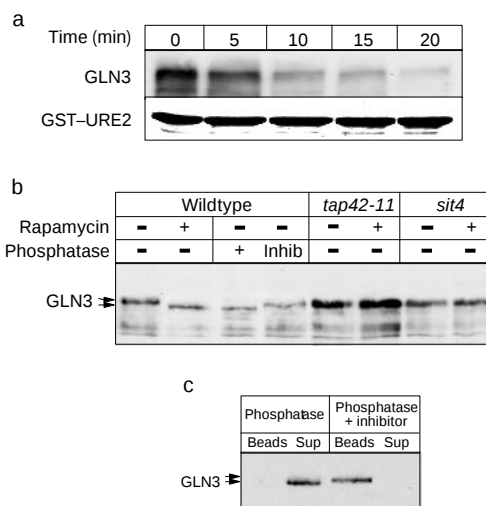


Figure 3 GLN3 and URE2 form a phosphorylation-dependent, rapamycin-sensitive complex. **a**, GLN3–URE2 complex is rapamycin sensitive. Equal amounts of GST–URE2 were purified from cells (TB123/pTB432) co-expressing GLN3–myc¹³ (GLN3) and treated with rapamycin for the indicated time before preparation of cell extracts. The amount of co-purified GLN3 was determined by immunoblotting. The total amount of GLN3 was unchanged throughout the rapamycin time course, and no unspecific binding of GLN3 to GST was detected (not shown). **b**, GLN3 is dephosphorylated in a SIT4-dependent manner upon rapamycin treatment. GLN3-myc¹³ (GLN3) expressed in wild-type (TB123), *tap42-11* (TB139-5b) or *sit4* (TB136-2a) cells untreated (–) or treated (+) for 10 min with rapamycin was detected by immunoblotting. In addition, extracts of rapamycin-untreated wild-type cells were incubated with alkaline phosphatase (+) or with alkaline phosphatase plus phosphatase inhibitors (Inhib). **c**, Phosphatase treatment releases GLN3 from URE2. GST–URE2 was purified from rapamycin-untreated cells (TB123/pTB432). Glutathione agarose beads with bound GST–URE2–GLN3 complex were resuspended in phosphatase buffer and treated with alkaline phosphatase in the absence or presence of phosphatase inhibitors. Supernatant (Sup) and beads were separated and analysed for GLN3 by immunoblotting.

occurred, accompanied by a strong decrease of cytoplasmic GLN3 (Fig. 2). Similarly, a nuclear accumulation of GAT1 was observed within 15–20 minutes of rapamycin treatment (data not shown). A cytoplasm-to-nucleus translocation of at least GLN3 was also observed in cells shifted from rich medium to medium containing a poor nitrogen source (proline) (data not shown), indicating that rapamycin-induced nuclear localization reflects a response to nitrogen limitation. Rapamycin did not induce nuclear accumulation of the GATA factors in a rapamycin-resistant *TOR-1* mutant¹⁴ (data not shown), indicating that the effect of rapamycin takes place through TOR. Thus, TOR appears to inactivate GLN3 and GAT1 by preventing their access to the nucleus.

How does TOR prevent nuclear localization of GLN3 and GAT1? URE2 represses GLN3-dependent transcription, possibly by a mechanism involving a direct interaction between GLN3 and URE2 (ref. 12). The above results indicate that URE2 may antagonize GLN3 by retaining the transcription factor in the cytoplasm. Consistent with this model, URE2 was considerably more abundant than GLN3, and localized in the cytoplasm in both untreated and rapamycin-treated cells (data not shown). To test whether URE2 acts by retaining GLN3 in the cytoplasm, we investigated the localization of GLN3 in cells lacking URE2 (*ure2*), and a possible URE2–GLN3 interaction. In contrast to wild-type cells, most GLN3 in *ure2* cells resided in the nucleus, even without rapamycin treatment, and a complete translocation into the nucleus was achieved upon a short exposure to rapamycin (Fig. 2). Furthermore, overexpression of URE2 largely prevented the rapamycin-induced translocation of GLN3 as observed in wild-type cells (data not

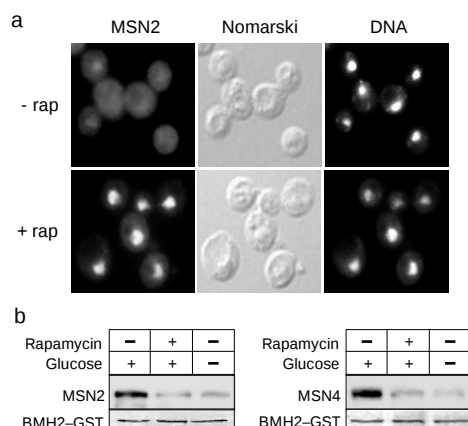


Figure 4 TOR promotes cytoplasmic retention of MSN and the BMH2-MSN interaction. **a**, MSN2 translocates to the nucleus upon TOR inactivation by rapamycin treatment. MSN2-myc (MSN2) was visualized by indirect immunofluorescence in wild-type cells untreated (-rap) or treated (+rap) with rapamycin for 20 min. Cells and DNA were visualized by Nomarski optics and DAPI staining (DNA). **b**, Binding of MSN to BMH2 is rapamycin sensitive and glucose dependent. Equal amounts of BMH2-GST were purified from cells co-expressing MSN2-myc (MSN2) or MSN4-HA³ (MSN4) grown in SD medium (see Methods). Before the preparation of extracts, cells were left untreated or were treated with rapamycin for 20 min; rapamycin-untreated cells were also shifted to glucose-free medium for 20 min. The amount of co-purified MSN was determined by immunoblotting. The total amounts of the MSN proteins were unchanged throughout the rapamycin time course, and no unspecific binding of MSN to GST was detected (not shown).

shown). To investigate a URE2-GLN3 interaction, a glutathione S-transferase (GST) 'pull-down' assay was carried out with whole-cell extracts prepared from cells co-expressing GST-URE2 and myc-tagged GLN3 (see Methods) (Fig. 3a). GLN3 co-purified with GST-URE2 from cells untreated with rapamycin. The URE2-GLN3 complex rapidly dissociated (within ~10 min) in cells treated with rapamycin. The kinetics of rapamycin-induced dissociation were similar to the kinetics of rapamycin-induced nuclear import of GLN3. A rapid dissociation of the complex was also observed upon transferring the cells from rich medium to medium containing a poor nitrogen source (proline) (data not shown), indicating that the rapamycin-induced dissociation indeed reflects a response to nutrient limitation. We did not detect significant complex formation between GAT1 and URE2 (data not shown). These results indicate that TOR prevents nuclear localization of at least GLN3 by tethering this GATA factor to URE2 in the cytoplasm.

How does TOR control URE2-GLN3 complex formation or stability? GLN3 and GAT1 in total cell extracts from rapamycin-treated cells exhibited higher electrophoretic mobility compared with these factors from untreated cells. No change was observed in the mobility of URE2 (Fig. 3b; and data not shown). Phosphatase treatment of whole-cell extracts from untreated cells converted the slower migrating forms of GLN3 and GAT1 to the faster migrating forms, indicating that the rapamycin-induced change in electrophoretic mobility may reflect a dephosphorylation of the transcription factors (Fig. 3b; and data not shown). TOR promotes complex formation between TAP42 and the type-2A-related phosphatase SIT4⁶, which possibly results in the inhibition of the phosphatase^{7,8}. To assess whether TAP42 and SIT4 are involved in the dephosphorylation of GLN3 and GAT1, we analysed the mobility of the transcription factors from strains containing the rapamycin-resistance *tap42-11* allele or a deletion of *SIT4* (*sit4*). Neither strain exhibited a rapamycin-induced bandshift (Fig. 3b; data not shown for GAT1), indicating that TOR controls the phosphorylation state of the GATA factors through TAP42 and SIT4. In addition, even after prolonged rapamycin treat-

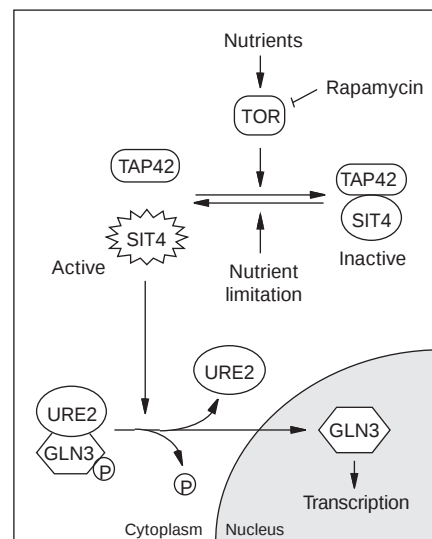


Figure 5 TOR promotes a cytoplasmic GLN3-URE2 complex via TAP42-mediated inhibition of the phosphatase SIT4, and thereby prevents nuclear accumulation of GLN3, in response to nutrients (nitrogen). TOR may similarly control an MSN-BMH complex and thereby prevent nuclear localization of MSN, in response to a good carbon source.

ment, only very weak nuclear accumulation of GLN3 and GAT1 was observed with the *tap42-11* strain, and no nuclear accumulation was detected with the *sit4* strain (Fig. 2; data not shown for GAT1). Closer examination of GST-URE2-bound GLN3 showed that only the slower migrating form of GLN3 was associated with URE2 (data not shown), indicating that GLN3 phosphorylation may be required for binding to URE2. In agreement with this possibility, *in vitro* dephosphorylation of the URE2-GLN3 complex resulted in a complete dissociation of GLN3 from GST-URE2 (Fig. 3c). Thus, TOR controls the URE2-GLN3 complex by maintaining GLN3 in a phosphorylated state necessary for URE2 binding. The results further indicate that TOR may maintain GLN3 phosphorylated by inhibiting SIT4 through TAP42.

Are GLN3 and GAT1 the only nutrient-regulated transcription factors controlled by TOR? MSN2 and MSN4 are partially redundant Zn²⁺-finger transcription factors that accumulate in the nucleus and activate a large number of stress-related genes in response to several types of stress including carbon limitation¹⁵⁻¹⁷. Genes activated by MSN2 and MSN4 (for example, *CTT1*, *HSP26* and *SSA3*) are induced upon rapamycin treatment². We examined the effect of rapamycin treatment on the cellular localization of MSN2 and MSN4. Similar to the GATA transcription factors, rapamycin treatment induced the nuclear accumulation of MSN2 and MSN4 within 20 minutes (Fig. 4a; and data not shown). Rapamycin failed to induce nuclear accumulation of at least MSN2 in a rapamycin-resistant *TOR1-1* mutant¹⁴ (data not shown).

14-3-3 proteins mediate nuclear exclusion of a forkhead transcription factor in mammalian cells in response to growth factors¹⁸. Furthermore, overexpression of 14-3-3 protein BMH1 or BMH2 confers rapamycin resistance in yeast¹⁹, similar to overexpression of URE2 (see above). To determine whether BMH2 might be retaining MSN (MSN2 and/or MSN4) in the cytoplasm, we investigated the cellular localization of BMH2, and the ability of BMH2 to bind MSN2 and MSN4. BMH2 was detected only in the cytoplasm in both untreated and rapamycin-treated cells, as determined by

immunofluorescence on whole fixed cells (data not shown). Both MSN2 and MSN4 bound to BMH2, as determined by a GST pull-down assay with whole-cell extracts from cells co-expressing BMH2–GST and a tagged MSN (Fig. 4b). Treatment of cells with rapamycin for 20 minutes before preparation of the whole-cell extract significantly reduced the co-purification of MSN in complex with BMH2.

To determine whether the interaction between BMH2 and MSN is carbon-source-regulated, as predicted by the rapamycin sensitivity of the complex (see above), we carried out the pull-down assay with cells shifted from glucose-containing medium to glucose-free medium 20 minutes before preparation of the whole-cell extract. A glucose shift was chosen because MSN-dependent transcription is induced in response to glucose withdrawal^{16,17}. Glucose withdrawal caused a release of MSN from BMH2, as seen previously with rapamycin treatment (Fig. 4b). The above results suggest that TOR, in response to a good carbon source, tethers MSN to BMH2 and thereby retains MSN in the cytoplasm. Contrary to what was observed for GLN3, the *tap42-11* and *sit4* mutations did not affect nuclear localization of MSN2 (data not shown), suggesting that TOR controls MSN by a different effector pathway. It will be of interest to determine the relationship between TOR and the RAS pathway in excluding MSN from the nucleus^{15–17}.

Our results indicate that TOR may prevent the nuclear accumulation of nutrient-regulated transcription factors (Fig. 5). The GLN3 and MSN transcription factors are retained in the cytoplasm by interaction with the abundant URE2 and 14-3-3 proteins, respectively. At least in the case of GLN3, phosphorylation of the transcription factor is required for this interaction. The phosphorylation state of GLN3 is controlled by TOR through TAP42-dependent inactivation of SIT4. It remains to be determined whether SIT4 dephosphorylates GLN3 directly. Interestingly, cytoplasmic retention and inactivation of a growth-inhibiting forkhead transcription factor by 14-3-3 proteins in mammalian cells is mediated by Akt/PKB^{18,20}. Thus, as mammalian TOR signals in conjugation with Akt/PKB and also controls growth in response to nutrients^{1,21}, our findings on the regulation of transcription by TOR may reflect conserved modes of TOR signalling and growth control. □

Methods

Strains, plasmids and media

All yeast strains used in this study are isogenic derivatives of JK9-3da (*MATa his4 leu2 ura3 trp1 rme1 HMLa*) and TB50a (*JK9-3da his3 HIS4*). Polymerase chain reaction (PCR)-based gene deletion (*gat1::HIS3MX*, TB101-1d; *gln3::HIS3MX*, TB103-1d; *gat1::HIS3MX gln3::kanMX*, TB105-3b) and tagging of chromosomal genes (*BMH2–GST–HIS3MX*, TB131; *GAT1–HA³–kanMX*, TB106-2a; *GLN3–myc¹³–kanMX*, TB123) were done as described^{22,23}. The *ure2* strain (*ure2::URA3, AS51-1a*) was kindly provided by A. Schmidt. The *tap42-11* strain (*tap42::kanMX HIS3 TRP1/YCplac111::tap42-11, TS54-5a*) and *sit4* strain (*sit4::kanMX, TS64-1a*) were kindly provided by T. Schmelzle. Haemagglutinin (HA)-tagged BMH2, MSN4 and URE2 were expressed under control of their own promoters from low copy number plasmids (*BMH2–HA³, pTB418*; *MSN4–HA³, pTB413*; *URE2–HA³, pTB377*) or from a high copy number plasmid (*URE2–HA³, pTB376*). *GST–URE2* was expressed under control of the *GAL* promoter from a low copy number plasmid (*pTB432*). The plasmid for expression of MSN2–myc was as described previously¹⁶. More detailed descriptions of strains and plasmids used in this study are available on request. Strains were grown in standard rich medium (YPD) or modified synthetic medium (SC) with either 2% glucose (SD) or 2% galactose (SGal) as the carbon source^{4,24}. Rapamycin (dissolved in 10% Tween-20/90% ethanol) was added to media at a final concentration of 200 ng ml⁻¹.

Northern blot analysis

Wild-type cells were grown to early logarithmic phase in rich medium at 30 °C. Culture aliquots were removed after 0, 10, 20 and 30 min of rapamycin treatment. Northern analysis was performed according to standard protocols²⁵. PCR products (~1 kilobase) were labelled using a DecaLabel DNA labelling kit (MPI Fermentas).

Phosphatase treatment

Treatment of whole-cell extracts was done as described⁷. Dephosphorylation of proteins bound to glutathione agarose beads was performed by incubation at 37 °C for 15 min, with 1 unit of alkaline phosphatase (Roche Diagnostics) in a volume of 30 µl in phosphatase buffer (supplied by the manufacturer). In control reactions, phosphatase inhibitors

(10 mM NaF, 10 mM *p*-nitrophenylphosphate, 10 mM sodium pyrophosphate, 10 mM β-glycerophosphate) were added.

GST pull-down

Cells expressing GST and GST–URE2 under control of the *GAL* promoter were grown to early logarithmic phase in SD medium. Expression of GST and GST–URE2 was induced by 4 h of growth in SGal, followed by 2 h in YPD medium. Cells expressing BMH2–GST were grown in SD medium to early logarithmic phase. Whole-cell extracts were prepared as described⁴. GST fusion proteins were bound to glutathione agarose beads (Pharmacia) and washed extensively with extraction buffer. Bound proteins were eluted with sample buffer (5 min, 95 °C). Ten per cent of each sample was subjected to immunoblotting, the remainder was subjected to SDS–PAGE and Coomassie-stained for protein quantification.

Miscellaneous methods

Immunoblotting and indirect immunofluorescence on fixed cells were done as described⁴. Preparation of whole-cell extracts was done as described⁴, with a modified extraction buffer (120 mM NaCl, 50 mM Tris pH 7.5, 2 mM EDTA, 2 mM PMSF, 10 mM NaF, 10 mM *p*-nitrophenylphosphate, 10 mM sodium pyrophosphate, 10 mM β-glycerophosphate). Cells were chilled briefly (~10 min) before centrifugation, for transfer to different nutrient conditions.

Received 15 October; accepted 11 November 1999.

1. Thomas, G. & Hall, M. N. TOR signalling and control of cell growth. *Curr. Opin. Cell Biol.* **9**, 782–787 (1997).
2. Barbet, N. C. *et al.* TOR controls translation initiation and early G1 progression in yeast. *Mol. Biol. Cell* **7**, 25–42 (1996).
3. Noda, T. & Ohsumi, Y. Tor, a phosphatidylinositol kinase homologue, controls autophagy in yeast. *J. Biol. Chem.* **273**, 3963–3966 (1998).
4. Beck, T., Schmidt, A. & Hall, M. N. Starvation induces vacuolar targeting and degradation of the tryptophan permease in yeast. *J. Cell Biol.* **146**, 1227–1237 (1999).
5. Dennis, P. B., Fumagalli, S. & Thomas, G. Target of rapamycin (TOR): balancing the opposing forces of protein synthesis and degradation. *Curr. Opin. Genet. Dev.* **9**, 49–54 (1999).
6. Di Como, C. J. & Arndt, K. T. Nutrients, via the Tor proteins, stimulate the association of Tap42 with type 2A phosphatases. *Genes Dev.* **10**, 1904–1916 (1996).
7. Schmidt, A., Beck, T., Koller, A., Kunz, J. & Hall, M. N. The TOR nutrient signalling pathway phosphorylates NPR1 and inhibits turnover of the tryptophan permease. *EMBO J.* **17**, 6924–6931 (1998).
8. Jiang, Y. & Broach, J. R. Tor proteins and protein phosphatase 2A reciprocally regulate Tap42 in controlling cell growth in yeast. *EMBO J.* **18**, 2782–2792 (1999).
9. Marzluf, G. A. Genetic regulation of nitrogen metabolism in the fungi. *Microbiol. Mol. Biol. Rev.* **61**, 17–32 (1997).
10. Marini, A. M., Soussi-Boudekou, S., Vissers, S. & Andre, B. A family of ammonium transporters in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **17**, 4282–4293 (1997).
11. Magasanik, B. In *The Molecular and Cellular Biology of the Yeast Saccharomyces* (eds Jones, E. W., Pringle, J. R. & Broach, J. R.) 283–317 (Cold Spring Harbor Laboratory Press, New York, 1992).
12. Blinder, D., Coschigano, P. W. & Magasanik, B. Interaction of the GATA factor Gln3p with the nitrogen regulator Ure2p in *Saccharomyces cerevisiae*. *J. Bacteriol.* **178**, 4734–4736 (1996).
13. Coschigano, P. W. & Magasanik, B. The URE2 gene product of *Saccharomyces cerevisiae* plays an important role in the cellular response to the nitrogen source and has homology to glutathione S-transferases. *Mol. Cell. Biol.* **11**, 822–832 (1991).
14. Heitman, J., Movva, N. R. & Hall, M. N. Targets for cell cycle arrest by the immunosuppressant rapamycin in yeast. *Science* **253**, 905–909 (1991).
15. Smith, A., Ward, M. P. & Garrett, S. Yeast PKA represses Msn2p/Msn4p-dependent gene expression to regulate growth, stress response and glycogen accumulation. *EMBO J.* **17**, 3556–3564 (1998).
16. Gorner, W. *et al.* Nuclear localization of the C2H2 zinc finger protein Msn2p is regulated by stress and protein kinase A activity. *Genes Dev.* **12**, 586–597 (1998).
17. Boy-Marcotte, E., Perrot, M., Bussereau, F., Boucherie, H. & Jacquet, M. Msn2p and Msn4p control a large number of genes induced at the diauxic transition which are repressed by cyclic AMP in *Saccharomyces cerevisiae*. *J. Bacteriol.* **180**, 1044–1052 (1998).
18. Brunet, A. *et al.* Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. *Cell* **96**, 857–868 (1999).
19. Bertram, P. G., Zeng, C., Thorson, J., Shaw, A. S. & Zheng, X. F. The 14-3-3 proteins positively regulate rapamycin-sensitive signaling. *Curr. Biol.* **8**, 1259–1267 (1998).
20. Kops, G. J. *et al.* Direct control of the Forkhead transcription factor AFX by protein kinase B. *Nature* **398**, 630–634 (1999).
21. Hara, K. *et al.* Amino acid sufficiency and mTOR regulate p70 S6 kinase and eIF-4E BP1 through a common effector mechanism. *J. Biol. Chem.* **273**, 14484–14494 (1998); erratum, 22160 (1998).
22. Wach, A., Brachat, A., Pohlmann, R. & Philippsen, P. New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. *Yeast* **10**, 1793–1808 (1994).
23. Longtine, M. S. *et al.* Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14**, 953–961 (1998).
24. Sherman, F. In *Guide to Yeast Genetics and Molecular Biology* (eds Guthrie, C. & Fink, G. R.) 3–21 (Academic, San Diego, 1991).
25. Ausubel, F. M. *et al.* (eds) *Current Protocols in Molecular Biology* (John Wiley & Sons, New York, 1998).

Acknowledgements

We thank T. W. Sturgill and T. Schmelzle for comments on the manuscript, E. Jacinto for unpublished results, and A. Löschmann for technical assistance. The work was supported by grants of the Swiss National Science Foundation and the Canton of Basel (M.N.H.).

Correspondence and requests for materials should be addressed to M.N.H. (e-mail: M.Hall@unibas.ch).

Eukaryotic type II chaperonin CCT interacts with actin through specific subunits

Oscar Llorca*, Elizabeth A. McCormack†, Gillian Hynes†, Julie Grantham†, Jacqueline Cordell†, José L. Carrascosa*, Keith R. Willison†, José J. Fernandez‡ & José M. Valpuesta*

* Centro Nacional de Biotecnología, C.S.I.C., Campus Universidad Autónoma de Madrid, 28049 Madrid, Spain

† CRC Centre for Cell and Molecular Biology, Institute of Cancer Research, Chester Beatty Laboratories, 237 Fulham Road, Chelsea, London SW3 6JB, UK

‡ Departamento de Arquitectura de Computadores y Electrónica, Universidad de Almería, 04120 Almería, Spain

Chaperonins assist the folding of other proteins¹. Type II chaperonins, such as chaperonin containing TCP-1 (CCT), are found in archaea and in the eukaryotic cytosol². They are hexadecameric or nonadecameric oligomers composed of one to eight different polypeptides. Whereas type I chaperonins like GroEL are promiscuous, assisting in the folding of many other proteins¹, only a small number of proteins, mainly actin and tubulin, have been described as natural substrates of CCT. This specificity may be related to the divergence of the eight CCT subunits³. Here we have obtained a three-dimensional reconstruction of the complex between CCT and α -actin by cryo-electron microscopy and image processing. This shows that α -actin interacts with the apical domains of either of two CCT subunits. Immunolabelling of CCT-substrate complexes with antibodies against two specific CCT subunits showed that actin binds to CCT using two specific and distinct interactions: the small domain of actin binds to CCT δ and the large domain to CCT β or CCT ϵ (both in position 1,4 with respect to δ). These results indicate that the binding of actin to CCT is both subunit-specific and geometry-dependent. Thus, the substrate recognition mechanism of eukaryotic CCT may differ from that of prokaryotic GroEL.

Chaperonins share an overall structure of two oligomeric rings, placed back-to-back. The atomic structure of the type I chaperonin GroEL from *Escherichia coli*⁴ and the type II chaperonin thermosome from *Thermoplasma acidophilum*⁵ have a similar organization in their three domains (apical, intermediate and equatorial) within the monomer. Type I chaperonins act with a cochaperonin, a small heptamer that seals the chaperonin cavity, whereas type II chaperonins seem to have a 'built-in' cochaperonin in the form of a helical protrusion^{5,6}. Whereas GroEL uses a mechanism based on non-specific hydrophobic interactions to interact with its substrate¹, the recognition of chemically denatured actin by CCT requires the generation of folding intermediates which form slowly after removal of the denaturant. This contrasts with the rapid binding of actin to GroEL⁷, indicating that actin binds to CCT in a conformation different from the one that is recognized by GroEL. Furthermore, GroEL cannot productively fold actin⁷. Actin appears to interact with CCT through defined regions⁸. Therefore, we have determined the structure of actin trapped on CCT to investigate the molecular recognition mechanism of type II chaperonins.

Native α -actin was chemically denatured and incubated with CCT in the absence of nucleotide, and aliquots of the solution were vitrified and observed at low temperature (-170°C) under an electron microscope. Images were taken at 20° tilt and digitized, and 3,546 top views were extracted and processed (see Methods and Supplementary Information). From the several classification procedures performed, two main groups of particles were obtained: those with a mass in the cavity (2,480 particles) and those with an empty cavity (766 particles). Both sets of particles were used for

independent three-dimensional reconstructions. The reconstruction generated from the CCT particles with empty cavities (Fig. 1a) shows a structure similar to that of nucleotide-free CCT⁹. The reconstruction generated with the substrate-bound particles reveals a structure with some differences (Fig. 1b–d): whereas one of the rings (the bottom one in Fig. 1b, c) is similar to either of the rings of the substrate-free CCT⁹, the other reveals the presence of a rod-shaped mass attached to the inner wall of the apical domains of two CCT subunits placed in a 1,4 arrangement (Fig. 1c, d).

The atomic structure of actin has a 'v' shape formed by two arms, the large and small domains¹⁰, each of which is divided into two subdomains. The structure of the CCT-bound α -actin shown here can be seen as the tips of the two open arms of α -actin interacting, respectively, with each of two CCT subunits. In addition, the dimensions of the α -actin reconstructed here ($\sim 90 \text{ \AA} \times 30 \text{ \AA}$) fit well with the dimensions of an opened-up conformation of the atomic structure of actin¹⁰ ($\sim 100 \text{ \AA} \times 35 \text{ \AA}$), so we surmise that α -actin could be substantially native-like before it interacts with CCT.

To corroborate this model of interaction, a chimaeric protein was constructed with subdomain 4 of the large domain of β -actin linked to the carboxy terminus of the Ha-Ras protein, Ha-Ras- β -actin subdomain 4 (called β -actin.sub4 from now on). This hybrid protein interacts with CCT (Fig. 2Ab and d) in reticulocyte lysate in equilibrium between a free, unbound population and a CCT-bound species; in contrast, β -actin is actively folded and processed to a non-CCT-interacting conformation in reticulocyte lysate (Fig. 2Aa and c). Recombinant β -actin.sub4 was chemically denatured and incubated with CCT. Images of the vitrified specimen were taken and digitized, and 3,361 20° -tilted top views of CCT were processed. Again, we obtained two sets of CCT particles after classification: 1,220 substrate-free particles and 2,030 particles with a mass inside the cavity. Two independent three-dimensional reconstructions were carried out. The one performed with the substrate-free particles gave rise to a similar structure to that shown in Fig. 1a (results not shown), whereas that generated using the substrate-bound particles revealed a mass bound to just one of the CCT subunits (Fig. 3), probably mediated through the actin subdomain of the chimaeric protein, as Ha-Ras does not interact with CCT (Fig. 2Ba and b).

An interesting conformational change takes place in the apical domains of the substrate-bound rings of both the CCT- α -actin and

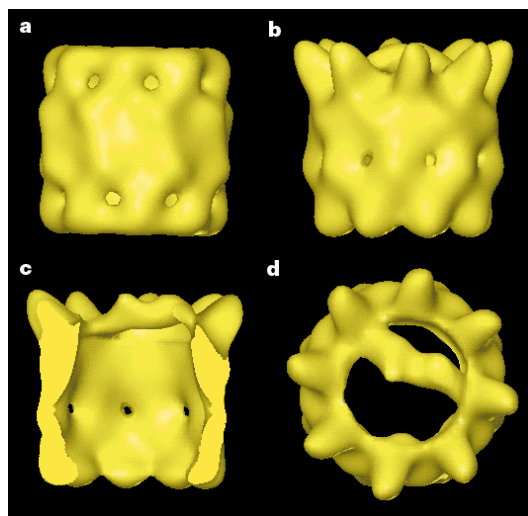


Figure 1 Three-dimensional reconstructions of CCT and the CCT- α -actin complex. **a**, Side view of the reconstruction obtained from substrate-free top views. **b–d**, Views of the three-dimensional reconstruction obtained from top views containing α -actin. **b**, Side view; **c**, a cut along the longitudinal axis of the CCT- α -actin complex; **d**, top view of the CCT- α -actin complex showing the actin-containing ring.

CCT- β -actin.sub4 three-dimensional reconstructions. The upwards and outwards movements of the apical domains induced upon substrate binding, either to one (for β -actin.sub4) or two CCT subunits (for α -actin), are probably important in the mechanism of protein folding by CCT and may also allow signalling between the

two CCT rings so that, once unfolded substrate binds to either of the rings, binding of a second polypeptide is prevented. CCT particles with substrate bound to both rings have not been observed despite the high (10:1) substrate:CCT molar ratio used.

The three-dimensional reconstruction of CCT- α -actin indicates

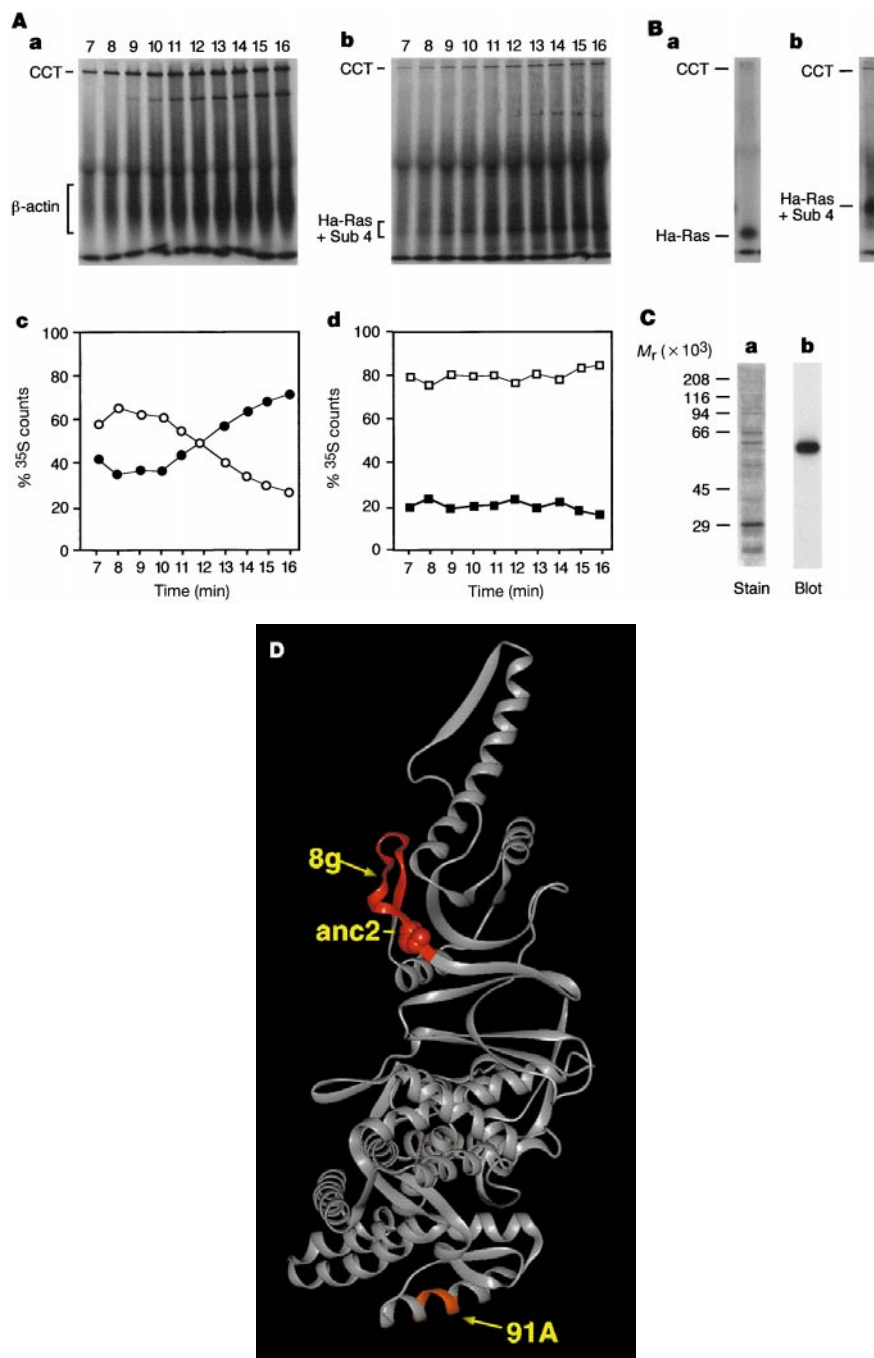


Figure 2 Interaction of β -actin and CCT. **A**, Analysis of the full-length β -actin (**a**, **c**) and β -actin.sub4 fusion protein (**b**, **d**) during an *in vitro* translation in rabbit reticulocyte lysate over 7–16-min. **a** and **b** show native polyacrylamide gel electrophoresis (PAGE) analysis of ^{35}S -methionine-labelled proteins (full-length actin contains 17 methionine residues and β -actin.sub4 only six). **c**, **d**, Quantitative analysis of the gel data shown in **a** and **b**, respectively, expressed as a percentage of the total ^{35}S -labelled β -actin (**c**) or β -actin.sub4 (**d**) bound (open symbols) or not bound (closed symbols) to CCT. **B**, Native PAGE analysis of *in vitro* translation at 60-min time points of Ha-Ras (**a**) and Ha-Ras + sub.4 (β -actin.sub4) (**b**) showing absence of interaction between Ha-Ras and CCT. **C**, Antibodies used for immunolocalization. The anti-CCT δ 8g monoclonal antibody used for electron microscope decoration is an IgG2b of high affinity which recognizes only a single polypeptide in western blots of total rabbit reticulocyte

lysate proteins. It does not recognize any other CCT subunit in western blot experiments using purified CCT (data not shown). **D**, Epitope localization. Solid-phase 15-mer peptide scanning arrays mapped the anti-CCT δ 8g epitope to between residues P344–S358 of CCT δ . The corresponding region is identifiably homologous in the thermosome⁵ and is located on the outer surface of the structure (shown here as a ribbon diagram of the α -thermosome subunit; PDB accession number 1A6D). The anti-CCT α 91A monoclonal antibody binds to the epitope A465–A469 (ref. 15), which is located in a clearly homologous region of the outer surface of the equatorial domain of the thermosome. The location of the yeast actin gene *anc2-1* mutation in the CCT4p/CCT δ subunit (residue G345D sequenced in this study (see Supplementary Information); *Saccharomyces cerevisiae* *anc2* strain was a gift from D. Drubin) is indicated as a space-filled residue in the equivalent glycine of the α -thermosome subunit.

that the tips of the arms of α -actin may interact with CCT through specific regions on the inner wall of the apical domains (apparently well below the helical protrusions^{5,6,9}) of two CCT subunits placed in a 1,4 arrangement. To determine whether this interaction is subunit-specific or only geometry-dependent, we used two monoclonal antibodies, the first against the apical domain of the CCT δ subunit (8g; Fig. 2C, D) and the second against the intermediate domain of the CCT α subunit (91A; Fig. 2D). The new anti-CCT δ monoclonal antibody was produced because a mutation in the yeast CCT4p/CCT δ subunit, *anc2*, rescues several mutations in subdomain 2 of yeast actin ACT1 (ref. 11 and Supplementary Information), indicating that this subunit may be involved in actin folding (Fig. 2D). Aliquots of the immunocomplexes were negatively stained (to contrast only one of the CCT rings) and images were recorded and digitized. We processed 970 untilted top views of the CCT- α -actin-8g complexes and obtained two main populations with substrate present in the CCT cavity (average images represented in Fig. 4a and b). In both average images, one

of the domains of actin interacts with CCT δ (the subunit to which the antibody binds) and the other with a subunit that is in position 1,4 with respect to CCT δ , either anticlockwise (Fig. 4a) or clockwise (Fig. 4b). According to the published model of CCT subunit arrangement¹² (Fig. 4j), the two CCT subunits involved in the interaction with actin would be the CCT β and CCT ϵ subunits. The processing of 1,069 untilted top views of CCT- β -actin.sub4-8g complexes again revealed two main populations. Their average images (Fig. 4c and d) show that subdomain 4 of β -actin binds to either of the two subunits placed in a 1,4 arrangement (clockwise or anticlockwise) to CCT δ (CCT β and CCT ϵ , respectively).

The results obtained with the 91A antibody reinforce those from the 8g antibody. Two populations were obtained from the processing of 2,054 untilted top views of CCT- α -actin-91A complexes (Fig. 4e and f). Antibody 91A binds to the CCT α subunit; the two average images generated show that one of the actin domains binds in both cases to a subunit that is in an anticlockwise 1,3 arrangement with respect to CCT α (CCT δ , according to ref. 12), and the other to

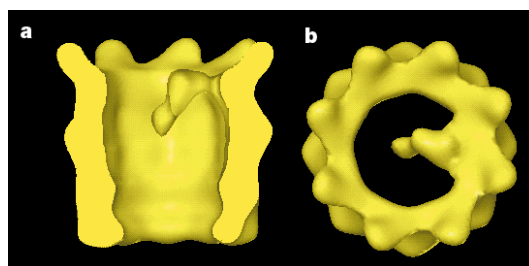


Figure 3 Three-dimensional reconstruction of the CCT- β -actin.sub4 complex. **a**, A cut along the longitudinal axis. **b**, Top view showing the β -actin.sub4-containing ring. The conditions are as described in Fig. 1.

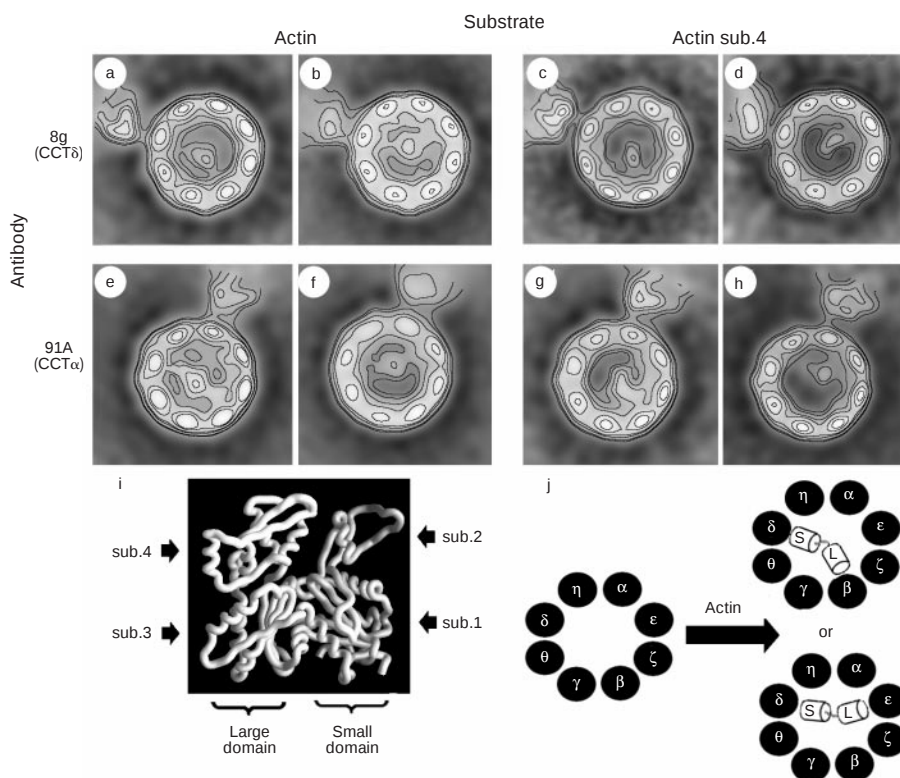


Figure 4 Two-dimensional average images of immunocomplexes of substrate-bound CCT. **a**, **b**, Two-dimensional average images of two classes of CCT- α -actin-8g complex. **c**, **d**, Two-dimensional average images of CCT- β -actin.sub4-8g complexes. **e**, **f**, Two-dimensional average images of CCT- α -actin-91A complexes. **g**, **h**, Two-dimensional average images of CCT- β -actin.sub4-91A complexes. No 8-fold symmetry

has been applied to any of these averages. **i**, Atomic structure of actin¹⁰ showing the large and small domains and their division into four subdomains (subs 1-4). **j**, The two possible modes of interaction of actin with CCT, using the CCT subunit orientation published in ref. 12, which is fully consistent with the results obtained here. L and S, large and small domains of actin, respectively¹⁰.

another CCT subunit that is in either a clockwise 1,2 or 1,4 position with respect to CCT α (Fig. 4e and f; CCT ϵ and CCT β subunits, respectively). The same two positions, CCT β and CCT ϵ , are occupied when two populations are generated out of 851 untilted top views of processed CCT- β -actin.sub4-91A complexes (Fig. 4g and h, respectively). The apparent paradox of obtaining a single three-dimensional reconstruction of the CCT- α -actin complex despite there being two possible ways for actin to interact with CCT results from the low resolution of this reconstruction, in which the two arms of actin and all of the CCT subunits are indistinguishable. Thus, the two possible three-dimensional reconstructions (with actin interacting with the CCT δ and CCT β subunits and with actin interacting with the CCT δ and CCT ϵ subunits) are the same if the first is rotated 135° with respect to the second (Fig. 1b-d).

Our results indicate that actin interacts with CCT by subunit-specific and geometry-dependent mechanism: the small domain of actin binds to the CCT δ subunit and the large domain can bind to either CCT β or CCT ϵ (see model in Fig. 4j). This type of interaction contrasts with the non-specific folding mechanism mediated by type I chaperonins like GroEL. Subjects for future study include the possible involvement of other CCT subunits in the interaction with tubulin³ and other substrates¹³, the conformational changes undergone by both substrate and CCT upon ATP binding and the hydrolysis needed to complete the CCT folding cycle. □

Methods

Antibodies

The anti-CCT δ 8g monoclonal antibody was prepared by immunizing rats with an apical domain fragment of the CCT δ subunit produced using recombinant *E. coli* (D219-N394 (6-His+23C epitope tagged at the C terminus) using a standard hybridoma protocol¹⁴. The anti-CCT α 91A monoclonal antibody was prepared as described^{14,15}.

Preparation of complexes

α -Actin (from rabbit muscle; Sigma) denatured in 7 M guanidinium chloride was incubated in a diluting buffer (100-fold) containing 0.4 μ M murine CCT purified as described¹². The Ha-Ras- β -actin fusion protein (β -actin.sub4) was constructed by linking residues 1-168 of human Ha-Ras to residues L178-F262 of human β -actin (subdomain 4). We prepared complexes of CCT- α -actin and CCT- β -actin.sub4 by incubation, for 15 min at room temperature, of CCT and either of the two substrates at a 1:10 molar ratio. In the case of the CCT-substrate-antibody complexes, preformed CCT- α -actin and CCT- β -actin.sub4 complexes were incubated with antibodies 8g or 91A (5:1 antibody: complex molar ratio) for 15 min at room temperature.

Electron microscopy

We applied aliquots of the immunocomplexes to carbon grids, negatively stained them with 2% uranyl acetate and recorded them at 0° tilt. Samples of the CCT-substrate solutions were applied to the grids for 1 min, blotted for 5 s and frozen quickly in liquid ethane at -180°C. Images were recorded at 20° tilt in a JEOL 1200EX-II electron microscope equipped with a Gatan cold stage operated at 120 kV and recorded on Kodak SO-163 film at a 60,000 $\times$ nominal magnification and ~2 μ m underfocus.

Image processing and three-dimensional reconstruction

In both negative stained and frozen-hydrated studies, top views were selected and two-dimensional processing was carried out as described⁹. In the case of the immunocomplexes, the average images obtained were filtered at the resolution obtained¹⁶ (22 Å). We used a self-organizing map algorithm¹⁷ to classify the particles into two main groups, those with no mass in the cavity of the particle and those with density in the cavity. With the two sets of particles, which have a large angular distribution, two independent three-dimensional reconstructions were performed using angular refinement algorithms provided by SPIDER¹⁸. The apo-CCT volume was used as the initial model for refinement, and the volumes were generated using ART with blobs (Algebraic Reconstruction Techniques¹⁹).

All of the two-dimensional processing carried out so far has revealed that the CCT particles have eightfold pseudosymmetry along the longitudinal axis^{9,16}. However, the particles containing substrate do not have eightfold symmetry in the cavity. For this reason, no symmetrization was applied to the inner cylinder comprising the CCT cavity. The final resolution was calculated by Fourier ring correlation of two independent reconstructions and the value obtained was used to low-pass filter the volumes (30 Å for the CCT- α -actin complex and 35 Å for the CCT- β -actin.sub4 complex).

Received 15 July; accepted 29 September 1999.

1. Bukau, B. & Horwich, A. L. The hsp70 and hsp60 chaperone machines. *Cell* **92**, 351-366 (1998).
2. Kubota, H., Hynes, G. & Willison, K. R. The chaperonin containing t-complex polypeptide 1 (TCP-1).

3. Multisubunit machinery assisting in protein folding and assembly in the eukaryotic cytosol. *Eur. J. Biochem.* **230**, 3-16 (1995).
4. Willison, K. R. in *Molecular Chaperones and Folding Catalysts* (ed. Bukau, B.) 555-571 (Harwood Academic, Amsterdam, 1999).
5. Braig, K. et al. The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature* **371**, 578-586 (1994).
6. Ditzel, L. et al. Crystal structure of the thermosome, the archaeal chaperonin and homologue of CCT. *Cell* **93**, 125-138 (1998).
7. Klumpp, M., Baumeister, W. & Essen, L. O. Structure of the substrate binding domain of the thermosome, an archaeal group II chaperonin. *Cell* **91**, 263-270 (1997).
8. Melki, R. & Cowan, N. J. Facilitated folding of actins and tubulins occurs via a nucleotide-dependent interaction between cytoplasmic chaperonin and distinctive folding intermediates. *Mol. Cell. Biol.* **14**, 2895-2904 (1994).
9. Rommelaere, H., De Neve, M., Melki, R., Vandekerckhove, J. & Ampe, C. The cytosolic class II chaperonin CCT recognizes delineated hydrophobic sequences in its target proteins. *Biochemistry* **38**, 3246-3257 (1999).
10. Llorca, O. et al. 3D reconstruction of the ATP-bound form of CCT reveals the asymmetric folding conformation of a type II chaperonin. *Nature Struct. Biol.* **6**, 639-642 (1999).
11. Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F. & Holmes, K. C. Atomic structure of the actin:DNase I complex. *Nature* **347**, 37-44 (1990).
12. Vinh, D. B.-N. & Drubin, D. G. A yeast TCP-1 like protein is required for actin function *in vivo*. *Proc. Natl Acad. Sci. USA* **91**, 9116-9120 (1994).
13. Liou, A. K. F. & Willison, K. R. Elucidation of the subunit orientation in CCT (chaperonin containing TCP1) from the subunit composition of micro-complexes. *EMBO J.* **16**, 4311-4316 (1997).
14. Kashuba, E., Pokrovskaja, K., Klein, G. & Szekeely, L. Epstein-Barr virus-encoded nuclear protein EBNA-3 interacts with the ϵ -subunit of the t-complex protein 1 chaperonin complex. *J. Hum. Virol.* **2**, 33-37 (1999).
15. Willison, K. R. et al. The t complex polypeptide 1 (TCP-1) is associated with the cytoplasmic aspect of Golgi membranes. *Cell* **57**, 621-632 (1989).
16. Hynes, G., Kubota, H. & Willison, K. R. Antibody characterization of two distinct conformations of the chaperonin containing TCP-1 from mouse testis. *FEBS Lett.* **358**, 129-132 (1995).
17. Llorca, O. et al. ATP binding induces large conformational changes in the apical and equatorial domains of the eukaryotic chaperonin containing TCP-1 complex. *J. Biol. Chem.* **273**, 10091-10094 (1998).
18. Marabini, R. & Carazo, J. M. Pattern recognition and classification of images of biological macromolecules using artificial neural networks. *Biophys. J.* **66**, 1804-1814 (1994).
19. Frank, J. et al. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190-199 (1996).
20. Marabini, R., Herman, G. T. & Carazo, J. M. 3D reconstruction in electron microscopy using ART with smooth spherically symmetric volume elements (blobs). *Ultramicroscopy* **72**, 53-65 (1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as hard copy from the London editorial office of Nature.

Acknowledgements

This work was partially supported by grants from the DGICYT (J.M.V. and J.L.C.). O.L. is a fellow the Comunidad Autónoma de Madrid. The UK laboratory is supported by the Cancer Research Campaign.

Correspondence and requests for materials should be addressed to J.M.V. (e-mail: jmv@cnb.uam.es).

Modified reaction centres oxidize tyrosine in reactions that mirror photosystem II

L. Kálmán^{*†}, R. LoBrutto[†], J. P. Allen^{*} & J. C. Williams^{*}

Center for the Study of Early Events in Photosynthesis and ^{*}Department of Chemistry and Biochemistry, and [†]Department of Plant Biology, Arizona State University, Tempe, Arizona 85287, USA

The participation of tyrosine in the oxidation of water by photosystem II, a multisubunit enzyme complex involved in plant photosynthesis, exemplifies the significant role amino-acid side chains play in oxidation/reduction reactions in proteins¹. The influence of the surrounding protein on the properties of amino-acid radicals and the attributes necessary for electron transfer are

[†] Permanent address: Institute of Biophysics, József Attila University, Szeged H-6722 Hungary.

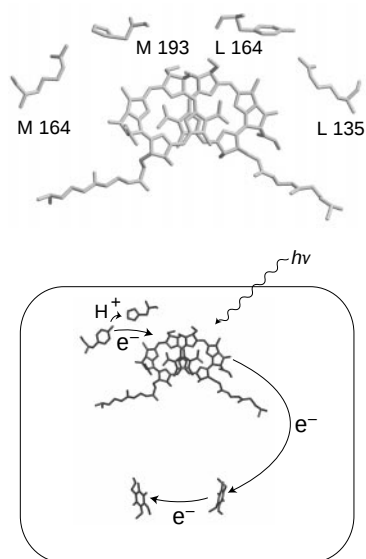


Figure 1 Model of bacterial reaction centre. Top, view of the bacteriochlorophyll dimer and nearby amino-acid residues in the reaction centre from *R. sphaeroides*²⁶. Tyrosine was substituted for Arg M164 and Arg L135 in the Y_M and Y_L reaction centres, respectively. These positions correspond to those of the Y_Z and Y_D tyrosine residues in photosystem II, and are approximately 10 Å from the edge of the bacteriochlorophyll dimer. Hydrogen bonds could be formed with His M193 in the Y_M mutant and with Tyr L164 in the Y_L mutant. Bottom, schematic drawing of the putative electron- and proton-transfer pathways in Y_M reaction centres.

not well understood. Here we report that modifications of reaction centres from the purple bacterium *Rhodospirillum rubrum* result in the generation of a tyrosyl radical in a manner similar to that of photosystem II. Our design incorporates a highly oxidizing bacteriochlorophyll dimer possessing a midpoint potential of more than 0.8 V, which was achieved by altering its local environment. After substitution of tyrosine residues at positions corresponding to the tyrosyl radicals of photosystem II, optical and electron paramagnetic resonance spectra showed changes consistent with oxidation of the tyrosine. The properties of this reaction include initiation by light and coupling to proton transfer, most probably to residues near the tyrosines.

Our design of a reaction centre capable of forming tyrosyl radical was based on the similarities between bacterial reaction centres and photosystem II (refs 2–4). Both of these pigment–protein complexes transfer an electron from an excited donor to quinone acceptors. In photosynthetic bacteria, the donor is subsequently reduced by a cytochrome in a cyclical process that forms an electrochemical proton gradient across the membrane. In contrast, photosystem II operates at significantly higher energies and generates potentials high enough to oxidize water. Consequently, modifications of the bacterial reaction centre were necessary to increase the oxidizing capacity of the bacteriochlorophyll dimer that constitutes the initial electron donor. Tyrosine residues substituted at sites that correspond to those of the Y_Z and Y_D radicals of photosystem II could then participate in new light-induced electron- and proton-transfer pathways (Fig. 1).

The midpoint potential of the dimer was raised incrementally from 0.5 V in the wild type to a value that would allow oxidation of tyrosine residues, whose midpoint potentials are expected to be at least 0.7 V. We raised the midpoint potential of the dimer by incorporating four substitutions in the reaction centre (Leu L131 to His, Leu M160 to His, Phe M197 to His and Tyr M210 to Trp). Each of these substitutions has been previously established to produce an increase of 0.06 V to 0.12 V in the midpoint potential^{5,6}. By performing oxidation–reduction titrations of the reaction

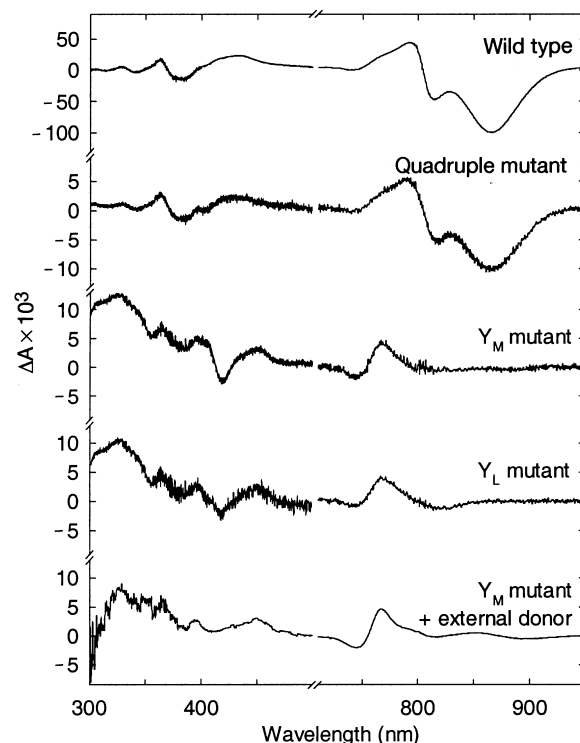


Figure 2 Light-minus-dark difference optical spectra. The quadruple mutant has similar spectral features with reduced amplitudes relative to the wild type, whereas the Y_M and Y_L mutants have distinct features indicative of a new state that accompanies the introduction of a tyrosine residue near the dimer. The features associated with reduced quinone in the wild type⁷ are seen in the spectrum of the Y_M mutant which has an added external electron donor, and in similar spectra of the other mutants.

centres from mutants with all four changes, we established a lower limit of 0.8 V for the midpoint potentials. Experimental limitations prevented us from measuring higher potentials. Spectroscopic measurements showed that the states formed by light-induced charge separation in the quadruple mutant, which contains only these four changes, are similar to wild type (Fig. 2). Based on the extent of the absorption changes we estimated the quantum yield of charge separation in these reaction centres to be 5–10%, compared with nearly 100% for wild type. The lower quantum yield results from a decrease in the free-energy difference between the excited and charge-separated states in reaction centres possessing elevated dimer midpoint potentials⁵. We found no evidence for specific oxidation of amino-acid residues in the quadruple mutant; this indicates that the higher potential alone is insufficient for radical formation.

Features indicative of a new state were observed in light-minus-dark difference optical spectra for reaction centres from the Y_M and Y_L mutants; in these, specific amino-acid residues were targeted for radical formation by substituting tyrosine at residues Arg M164 and Arg L135 (Fig. 2). In the near-infrared region, both the wild type and quadruple mutant have spectral features characteristic of the formation of an oxidized dimer and a reduced quinone, including a bleaching centred at 865 nm and overlapping electrochromic shifts centred near 760 and 800 nm. In contrast, in the spectra for the Y_M and Y_L reaction centres, there were no absorption changes above 800 nm associated with the oxidized dimer, but the spectral shift observed near 760 nm associated with the reduced quinone was still observed. The spectra in this region look similar to those of samples in which an external electron donor has been added to remove the contribution of the oxidized dimer⁷. The appearance of the reduced quinone but not the oxidized dimer indicates the presence of a kinetically viable electron donor for the light-oxidized dimer. In the

300–500-nm region, we observed a broad absorption increase centred near 430 nm and an electrochromic shift in the Soret band of the tetrapyrroles centred around 370 nm for both the wild type and the quadruple mutant. Spectral features of the Y_M and Y_L reaction centres include a derivative shape with a notable absorption decrease near 420 nm and an absorption increase near 300 nm, neither of which is associated with reduced quinone. Similar spectral features near 420 and 300 nm accompany the tyrosyl radicals in photosystem II (ref. 8).

We observed signals characteristic of tyrosyl radicals in room-temperature electron paramagnetic resonance (EPR) spectra of reaction centres from the Y_M and Y_L mutants, confirming the identity of the state observed in the optical spectra (Fig. 3). In the EPR spectra, reaction centres for both the wild type and the

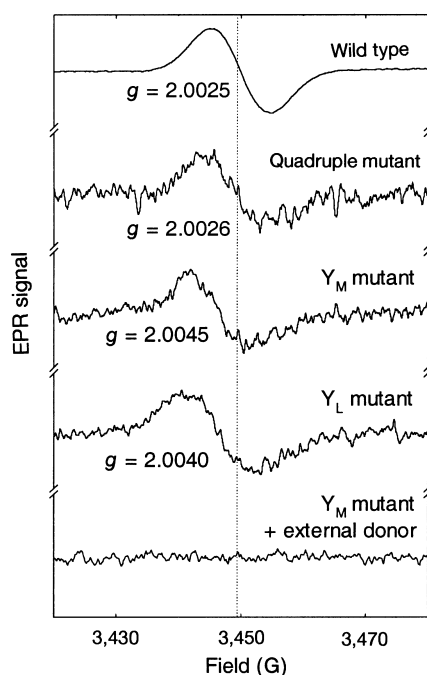


Figure 3 Light-minus-dark difference EPR spectra. The signals from the Y_M and Y_L mutants are shifted relative to those of the quadruple mutant and wild type, indicating the presence of a new radical species, most probably a tyrosyl radical. Although reduced quinones have similar g values, they do not contribute to these spectra, as can be seen when an external electron donor is added to remove the oxidized species in the Y_M mutant. Similarly, no quinone acceptor signal is observed in this region in reaction centres from the quadruple and Y_L mutants.

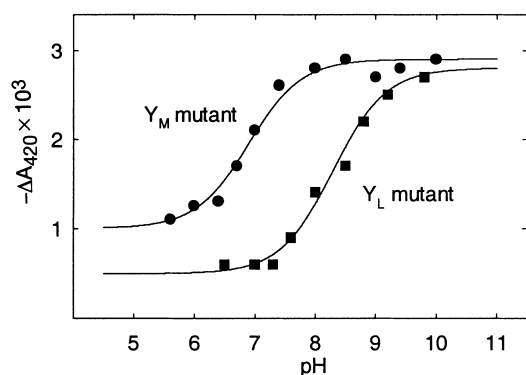


Figure 4 Dependence on pH of the light-induced negative absorption change at 420 nm. Fits yielded pK values of 6.9 and 8.3 for the Y_M and Y_L reaction centres, respectively, consistent with titration of nearby proton-accepting residues (Fig. 1).

quadruple mutant show the derivative signal for the oxidized dimer that is centred at $g = 2.0025$ (Ref. 9). In contrast, the spectra of the Y_M and Y_L reaction centres are clearly shifted to $g = 2.0045$ and 2.0040, respectively, with peak-to-trough linewidths of 11–12 G. Such a signal cannot be due to a bacteriochlorophyll and must arise from either a tyrosyl radical or a reduced quinone. In bacterial reaction centres, the quinones are magnetically coupled to the non-haem iron, resulting in a very broad signal¹⁰ and not the narrow signals observed for the Y_M and Y_L reaction centres. Reaction centres from the mutants showed no narrow signal when an external donor was present; this eliminates the possibility that an uncoupled quinone acceptor contributes to these spectra. Furthermore, experiments performed at 4 K revealed the typical broad quinone EPR spectrum in the quadruple and tyrosine mutants, confirming that the iron–quinone coupling is undisturbed in these samples (data not shown). A variety of spectral line shapes are observed for tyrosyl radicals in proteins and model systems^{11–13}. In photosystem II, signals with g values of 2.0046 and widths of 17–20 G are typically observed for the tyrosyl radicals². Linewidths of 10–11 G are found in mutants where nearby residues are altered^{11,14–16}, although a high-field EPR study of one such mutant showed that the signal has contributions from both tyrosyl and chlorophyll radicals¹⁷. We conclude that the signals measured for the Y_M and Y_L mutants originate predominantly from tyrosyl radicals, with the linewidths and g values decreased by a small ($\sim 10\%$) contribution from other radicals, most probably the oxidized dimer.

We demonstrated the influence of the local environment on the observed electron transfer events by comparing the pH dependence of the signals in the Y_M and Y_L reaction centres. The extent of the absorption decrease at 420 nm increased systematically with increasing pH in both mutants (Fig. 4). The other optical and EPR spectral features followed a similar reversible pattern with respect to pH, and at low pH values the reaction centres showed features of both the oxidized dimer and the tyrosyl radical. The pH dependence is explained by models of the mechanism of tyrosine oxidation in proteins, in which release of the phenolic proton to a nearby proton acceptor results in a neutral tyrosyl radical¹⁸. The pK value of 6.9 for the Y_M reaction centre is consistent with His M193 serving as a proton acceptor, similar to the role proposed for an analogous histidine residue in photosystem II (refs 19–24). The radical signal remaining at lower pH values indicates that another nearby residue, such as Glu M173, could also be involved in this type of interaction. The putative proton acceptor in the Y_L mutant is at Tyr L164 rather than histidine (Fig. 1), and the higher pK value of 8.3 can be attributed to the difference in the protonation states of the proton-accepting residues. The tyrosyl-radical signals in photosystem II are also pH dependent, with a shift in pK observed when tyrosine is introduced as a proton acceptor for Y_Z (ref. 22).

Both the optical and EPR data clearly show that the tyrosine residues introduced at M164 and L135 can readily donate an electron to the light-oxidized bacteriochlorophyll dimer. Thus, a specific electron-transfer event occurs in the Y_L and Y_M mutants rather than nonspecific oxidation caused by the presence of the highly oxidizing dimer. The ability to introduce tyrosine residues that can be selectively oxidized after excitation by light demonstrates that functional electron-transfer cofactors can be designed and created in biological systems. Although the changes made in these bacterial reaction centres do not duplicate the protein sequence or pigments found in photosystem II, the modifications to the environment of the bacteriochlorophyll dimer that we describe make it energetically capable of oxidizing water, modelling a property that has been proposed for an evolutionary intermediate between ancestral reaction centres and photosystem II (ref. 25). Other properties are needed for efficient water oxidation. For example, because the initial electron transfer proceeds from the excited state of the dimer, substitution of chlorophyll for bacteriochlorophyll would increase the driving force for the forward electron transfer

through an additional 0.5 eV of excitation energy, restoring the high quantum yield. The addition of a manganese complex in photo-system II provides the required storage capacity for the four electron equivalents used in the production of molecular oxygen. □

Methods

Construction of mutants

Mutations were made in *R. sphaeroides* using procedures described previously⁵. The data reported are from three reaction-centre mutants: the quadruple mutant LH(L131)+LH(M160)+FH(M197)+YW(M210), the Y_M mutant LH(L131)+LH(M160)+RY(M164)+FH(M197)+YW(M210) and the Y_L mutant LH(L131)+RY(L135)+LH(M160)+FH(M197)+YW(M210). All three mutants contain the substitutions Leu L131 to His, Leu M160 to His, Phe M197 to His and Tyr M210 to Trp, each of which results in an increase in the midpoint potential of the dimer. The Y_M and Y_L mutants contain the additional changes Arg M164 to Tyr and Arg L135 to Tyr, respectively.

Optical and EPR spectra

The spectra of isolated reaction centres were taken at room temperature using a Varian Cary 5 spectrophotometer and a Bruker E580 X-band spectrometer. Illumination was performed with an Oriol tungsten lamp, using low light levels that resulted in only 30% of charge separation in wild type and fully reversible spectral changes in the mutants. Samples were in 0.05% Triton X-100, 100 mM sodium chloride and 100 µM terbutryn, with a protein concentration of 1.5 µM for the optical spectra and 300 µM for the EPR spectra. Buffers used were 15 mM Tris-HCl, pH 8.0, or 7.5 mM Tris-HCl and 7.5 mM Ches, pH 9.0. In the spectra shown, the reaction centres from the Y_L mutant were at pH 9, compared with pH 8 for the other samples. At pH 9, the tyrosine features predominate in this mutant, whereas at pH 8 the oxidized dimer makes a larger contribution. The spectra of the reduced quinone were obtained after addition of 100 µM diaminodurene for the optical spectra and 3 mM cytochrome *c* for the EPR spectra. The *g* values indicated in the EPR spectra include corrections for small frequency differences in the measurements and have estimated errors of ±0.0005. The microwave frequency was 9.66 GHz, with a microwave power of 10 mW. A liquid-helium gas-flow cryostat was used to measure EPR spectra at 4 K.

Titration

Electrochemical oxidation–reduction titrations of isolated reaction centres were performed as described previously⁵. Above an ambient potential of approximately 0.8 V, significant degradation of the samples occurred, precluding measurements in this regime. For the pH titrations, the buffering capacity was provided by a combination of Mes, Tris and Ches buffers. The data from the pH titrations were fitted to a Henderson–Hasselbalch equation.

Received 4 June; accepted 21 October 1999.

1. Sigel, H. & Sigel, A. (eds) *Metalloenzymes Involving Amino-Acid Residue and Related Radicals* (Dekker, New York, 1994).
2. Diner, B. A. & Babcock, G. T. in *Oxygenic Photosynthesis* (eds Ort, D. R. & Yocum, C. F.) 213–247 (Kluwer, Dordrecht, 1996).
3. Rhee, K.-H. *et al.* Two-dimensional structure of plant photosystem II at 8-Å resolution. *Nature* **389**, 522–526 (1997).
4. Allen, J. P. & Williams, J. C. Photosynthetic reaction centers. *FEBS Lett.* **438**, 5–9 (1998).
5. Lin, X. *et al.* Specific alteration of the oxidation potential of the electron donor in reaction centers from *Rhodospirillum rubrum*. *Proc. Natl Acad. Sci. USA* **91**, 10265–10269 (1994).

6. Nagarajan, V., Parson, W. W., Davis, D. & Schenck, C. C. Kinetics and free energy gaps of electron-transfer reactions in *Rhodospirillum rubrum* reaction centers. *Biochemistry* **32**, 12324–12336 (1993).
7. Verméglio, A. & Clayton, R. K. Kinetics of electron transfer between the primary and the secondary electron acceptor in reaction centers from *Rhodospirillum rubrum*. *Biochim. Biophys. Acta* **461**, 159–165 (1977).
8. Diner, B. A. *et al.* in *Photosynthesis: From Light to Biosphere* Vol. II (ed. Mathis, P.) 229–234 (Kluwer, Dordrecht, 1995).
9. McElroy, J. D., Feher, G. & Mauzerall, D. C. Characterization of primary reactants in bacterial photosynthesis. *Biochim. Biophys. Acta* **267**, 363–374 (1972).
10. Isaacson, R. A., Lendzian, F., Abresch, E. C., Lubitz, W. & Feher, G. Electronic structure of Q_A in reaction centers from *Rhodospirillum rubrum*. *Biophys. J.* **69**, 311–322 (1995).
11. Tang, X.-S., Chisholm, D. A., Dismukes, G. C., Brudvig, G. W. & Diner, B. A. Spectroscopic evidence from site-directed mutants of *Synechocystis* PCC6803 in favor of a close interaction between histidine 189 and redox-active tyrosine 160, both of polypeptide D2 of the photosystem II reaction center. *Biochemistry* **32**, 13742–13748 (1993).
12. Hoganson, C. W., Sahlin, M., Sjöberg, B.-M. & Babcock, G. T. Electron magnetic resonance of the tyrosyl radical in ribonucleotide reductase from *Escherichia coli*. *J. Am. Chem. Soc.* **118**, 4672–4679 (1996).
13. Stubbe, J. & van der Donk, W. A. Protein radicals in enzyme catalysis. *Chem. Rev.* **98**, 705–762 (1998).
14. Tommos, C. *et al.* Modified EPR spectra of the tyrosine_D radical in photosystem II in site-directed mutants of *Synechocystis* sp. PCC 6803: identification of side chains in the immediate vicinity of tyrosine_D on the D2 protein. *Biochemistry* **32**, 5436–5441 (1993).
15. Roffey, R. A., van Wijk, K. J., Sayre, T. S. & Stryer, S. Spectroscopic characterization of tyrosine-Z in histidine 190 mutants of the D1 protein in photosystem II (PSII) in *Chlamydomonas reinhardtii*. *J. Biol. Chem.* **269**, 5115–5121 (1994).
16. Manna, P., LoBrutto, R., Eijkelhof, C., Dekker, J. P. & Vermaas, W. Role of Arg180 of the D2 protein in photosystem II structure and function. *Eur. J. Biochem.* **251**, 142–154 (1998).
17. Un, S., Tang, X. S. & Diner, B. A. 245 GHz high-field EPR study of tyrosine-D[•] and tyrosine-Z[•] in mutants of photosystem II. *Biochemistry* **35**, 679–684 (1996).
18. Hoganson, C. W. & Babcock, G. T. A metal-radical mechanism for the generation of oxygen from water in photosynthesis. *Science* **277**, 1953–1956 (1997).
19. Gilchrist, M. L., Jr, Ball, J. A., Randall, D. W. & Britt, R. D. Proximity of the manganese cluster of photosystem II to the redox-active tyrosine Y_Z. *Proc. Natl Acad. Sci. USA* **92**, 9545–9549 (1995).
20. Hoganson, C. W. *et al.* A hydrogen-atom abstraction model for the function of Y_Z in photosynthetic oxygen evolution. *Photosynthesis Res.* **46**, 177–184 (1995).
21. Tang, X.-S., Randall, D. W., Force, D. A., Diner, B. A. & Britt, R. D. Manganese-tyrosine interaction in the photosystem II oxygen-evolving complex. *J. Am. Chem. Soc.* **118**, 7638–7639 (1996).
22. Mamedov, F., Sayre, R. T. & Stryer, S. Involvement of Histidine 190 on the D1 protein in electron/proton transfer reactions on the donor side of photosystem II. *Biochemistry* **37**, 14245–14256 (1998).
23. Hays, A.-M. A., Vassiliev, I. R., Golbeck, J. H. & Debus, R. J. Role of D1-His190 in proton-coupled electron transfer reactions in photosystem II: a chemical complementation study. *Biochemistry* **37**, 11332–11365 (1998).
24. Ahlbrink, R. *et al.* Function of tyrosine Z in water oxidation by photosystem II: electrostatic promoter instead of hydrogen abstractor. *Biochemistry* **37**, 1131–1142 (1998).
25. Blankenship, R. E. & Hartman, H. The origin and evolution of oxygenic photosynthesis. *Trends Biol. Sci.* **23**, 94–97 (1998).
26. Feher, G., Allen, J. P., Okamura, M. Y. & Rees, D. C. Structure and function of bacterial photosynthetic reaction centres. *Nature* **339**, 111–116 (1989).

Acknowledgements

We thank X. Nguyen, C. Magee, J. McCulley and N. Boruck for their assistance with this project, and W. Lubitz and F. Lendzian for preliminary EPR measurements. This work was supported by grants from the NSF and USDA.

Correspondence and requests for materials should be addressed to J.P.A. (e-mail: jallen@asu.edu).

Cell biologicals

The last New on the Market before Christmas... but no last-minute gift ideas.

Direct-Q water purifier

Millipore www.millipore.com

Ultrapure (Type 1) water direct from tap feed

The Direct-Q Laboratory Water Purification System is designed for scientists who need 5 to 15 litres of ultrapure water per day and do not have access to acceptable quality pre-treated water. This compact system produces ultrapure water at up to 0.6 litres per minute, direct from potable tap water, without the need for pre-treatment. The water is suitable for most lab applications including chromatography and in culture media. Direct-Q can be bench-top operated or wall-mounted.

Reader Service No. 100

ApoAlert

Clontech Laboratories www.clontech.com

An assay to detect an early apoptotic event

The ApoAlert Mitochondrial Membrane Sensor Kit is a simple fluorescence-based method for distinguishing between healthy and apoptotic cells by detecting changes in mitochondrial membrane potential. The kit uses the cationic dye MitoSensor. In healthy cells, MitoSensor is taken up in the mitochondria, forming aggregates with intense red fluorescence. In apoptotic cells, the dye cannot aggregate in the mitochondria because of the altered mitochondrial transmembrane potential, so it remains in monomeric form in the cytoplasm, where it fluoresces green.

Reader Service No. 101

DePsipher

Trevigen www.trevigen.com

Measure the collapse of mitochondrial electrochemical potential during apoptosis.

DePsipher (pronounced de-SY-fer) is a lipophilic cationic dye that forms red-orange aggregates within healthy mitochondria. Following induction of apoptosis, pores form in the mitochondrial membrane causing a loss of membrane potential. In these apoptotic cells, the DePsipher reagent remains in the cytoplasm as a green fluorescent monomer.

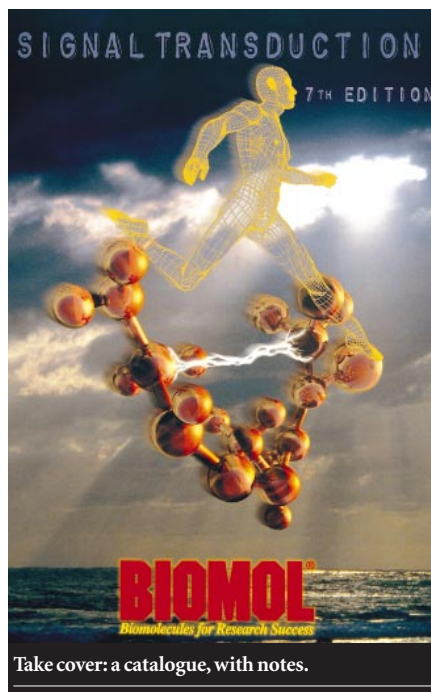
Reader Service No. 102

Signal transduction reagents

Biomol www.biomol.com

Reference sections and new products in a new catalogue

Biomol's 7th Edition Catalog/Handbook lists more than 1,300 signal transduction products including bioactive lipids and neuropharmacology. Sections expanded in this edition include those on apoptosis, nitric oxide and



protein phosphorylation research, current literature references and application notes.

Reader Service No. 103

Fas Ligand ELISA

Oncogene Research Products

www.oncresprod.com

An assay for both forms of FasL

This Fas Ligand ELISA is suitable for the quantification of soluble and membrane bound fas ligand in cell lysates, tissue culture medium and human sera and plasma. It is a colorimetric assay that can be completed in 5 hours with a sensitivity of 0.013 ng per ml and an assay range 0.3125–5 ng per ml. For more information visit the company's web site or www.apoptosis.com

Reader Service No. 104

Monoclonal antibodies

InnoGenex www.innogenex.com

Antibodies to cell cycle associated proteins

Monoclonal antibody E65.1 reacts with the 59-kD cyclin A protein of human, murine, simian, and bovine origin. Clone POH-1 recognizes the p34cdc2 cyclin-dependent kinase, identified as the catalytic subunit of the maturation-promoting factor. Monoclonal antibody EA10 was raised against recombinant human p21 cdk inhibitor. Also available is a purified rabbit polyclonal IgG raised against a synthetic peptide encompassing amino acids 130–164 of human p21

cdk inhibitor. A rabbit polyclonal against p16INK4A cdk inhibitor is also offered.

Reader Service No. 105

IPO-4 antibody

Kamiya Biomedical

www.kamiyabiomedical.com

A monoclonal to induce apoptosis

This new IgM monoclonal antibody (clone IPO-4) induces apoptosis in Fas expressing human cells. IPO-4 is an "excellent killer" that does not require the addition of a secondary cross-linker. IPO-4 can also be used for western blots, immunohistology, flow cytometry, and fluorescence microscopy.

Reader Service No. 106

Antibody to TRAIL

Upstate Biotechnology

www.upstatebiotech.com

Antibody to human TRAIL

TRAIL is a type II transmembrane glycoprotein, from the TNF cytokine family, that functions in a similar way to Fas ligand and induces apoptosis in many transformed cell lines, but not normal tissues. This new antibody, raised against amino acids 261–277 of human TRAIL, is supplied as 100 µg of protein G purified rabbit IgG in 200 µl PBS.

Reader Service No. 107

Chemokine ligands

NEN Life Science www.nenlifesci.com

Four new radiolabelled chemokine ligands

These chemokine ligands are intended for disease-focused receptor research. [125I]MIP-3α selectively binds to CCR6. [125I] MIP-3β selectively binds to CCR7. [125I]TARC selectively binds to CCR4 And [125I]Fractalkine selectively binds to CX3CR1. These introductions bring the NEN range to more than 20 chemokine ligands.

Reader Service No. 108



Read the label: NEN's new ligands.

Leica DM IL HC

Leica Microsystems www.leica-microsystems.com

An inverted microscope for live cell research

The DM IL HC is designed for brightfield, phase contrast, incident light fluorescence and high-quality relief contrast can be produced with the Integrated Modulation Contrast (IMC) technique. This gives contrast without the need for special objectives, giving a high-contrast 3D image of transparent objects similar to that of interference contrast. **Reader Service No. 109**

SignalScreen

Packard BioScience www.packardinst.com

A new range of G-protein coupled receptors

The SignalScreen range, developed by Packard's BioSignal subsidiary, comprises over 60 cloned G Protein Coupled Receptors. SignalScreen cloned receptors are purified and ready for membrane preparation for customers interested in using them in a pharmaceutical screening program. **Reader Service No. 110**

Na/H exchange antibodies

Chemicon www.chemicon.com

Antibodies to Na⁺/H⁺ exchangers

Na⁺/H⁺ exchange is involved in many cellular functions including cell volume and pH regulation. In mammalian cells a gene family consisting of at least five isoforms, NHE1-5, mediates the exchange. Chemicon have now introduced monoclonal antibodies to NHE1 and NHE3. Rabbit and chicken polyclonal antibodies to NHE1-5 are also available. **Reader Service No. 111**

IntraStain kit

Alexis Biochemicals www.al Alexis-corp.com

Intracellular cytokine detection

Studies of cell cycle analysis, cytokine production or oncogene expression often require analysis of cytoplasmic and/or nuclear anti-



Leica this: live cells under the microscope.

gens, necessitating permeabilization of the cell membrane to allow interaction of the antibody or probe with its target. Alexis Biochemicals produce the IntraStain Kit for this purpose. The four-reagent kit blocks Golgi function, stabilizes cellular integrity, permeabilizes the cellular membrane, washes and resuspends cells for analysis. **Reader Service No. 112**

Growth factor assay

Biosource www.lifescrreen.com

A human hepatocyte growth factor ELISA

The Biosource human HGF ELISA kit is rapid, completed in about 3 hours, with a high sensitivity (10 pg), and a wide range (156-10,000 pg per ml). It is useful for investigating HGF levels in serum, plasma or cell culture supernate and is available as 1 or 2 × 96 well plate kits. **Reader Service No. 113**

TotalLab

NonLinear Dynamics www.totallab.com

Multi-function software for the laboratory

TotalLab has facilities for image capture, the

documentation and analysis of 1D and 2D gels, dot and slot blots, microtitre plates, other gridded arrays and colonies. Features include automatic analysis of 1D gels in under 5 seconds for densitometry and molecular sizing, full quantitative and qualitative analysis of arrays, and colony counting based on very powerful 2D spot detection algorithms. Fully working demos of the software, with a large number of sample images for testing, can be downloaded from the TotalLab web site. **Reader Service No. 114**

Xplore

Endogen, Inc. www.endogen.com

Assays for the quantitative measurement of mouse GAPDH, IFN-γ and IL-1-β mRNA

Xplore Assays provide specific measurement of mouse cytokine gene expression in five hours using a convenient microtitre plate format and fluorescence detection. Xplore assays measure mRNA levels directly from samples without PCR amplification or radioactivity. The gel-free procedure requires just 0.5 μg of total RNA or poly A⁺ mRNA sample. **Reader Service No. 115**

BX40Cy microscope

Olympus www.olympus-europa.com

Are you sitting comfortably?

BX Series microscopes feature adjustable eyepoint making them particular suitable for cyto-screener and other users who spend the majority of their time examining slide samples. These microscopes can accommodate up to two 30 mm extension rings inserted between the microscope frame and the observation tube. Cyto-screener can also incline the observation tube between 5° and 35° which means they do not have to adopt an unnatural posture. **Reader Service No. 116**

These notes are compiled in the Nature office from information provided by the manufacturers. For more details, fill in the reader service card.

ADVERTISEMENTS

Homogenizers?

- Tissue-Tearor™
- BeadBeater™
- Mini-BeadBeater™
- BioHomogenizer™

Check Us Out

www.biospec.com

BioSpec Products, Inc.
Ph/Fax 918-336-3363
e-mail info@biospec.com

New Probe Sizes
0.45cm!

Customised monoclonal & polyclonal antibodies with guarantee!

BioGenes GmbH
Köpenicker Str. 325 · D-12555 Berlin
Tel. +49-30-65 76 23 96 Fax: - 97
E-mail: BioGenes@Berlin.sireco.net
Internet: <http://www.biogenes.de>

Check out our homepage at <http://www.resgen.com>

Peptides
\$14/residue
10 residue minimum

Polyclonal Antibodies - \$1,195

Please contact us about complete services for your custom peptide and polyclonal antibody needs.

Research Genetics, Inc.
U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 256-536-9016